

Kolmogorov-Arnold Networks for Precipitation Forecasting in Switzerland

Elwin Freudiger

Abstract.

Key words.

1. Literature Review.

1.1. Weather Forecasting. Weather forecasting is a critical tool across multiple sectors, from agriculture and transportation to tourism and emergency response. While the idea of predicting atmospheric states dates back to ancient civilizations, scientific methods have only been systematically applied since the nineteenth century [65].

In recent years, the field has undergone a profound transformation through the integration of traditional numerical weather prediction (NWP) techniques with modern machine learning (ML) approaches. These developments have improved both the accuracy and resolution of forecasts, especially in areas with complex topography, such as the Alps [59].

Numerical Weather Prediction. Numerical Weather Prediction remains a cornerstone of meteorology and weather forecasting. First proposed by Abbe and Bjerknes, the approach is based on the idea that physical laws governing the atmosphere can be used to predict future conditions. This involves solving systems of equations that describe atmospheric and oceanic dynamics [7].

Manual NWP was first attempted in 1922 in Britain, while computer-aided predictions emerged in the 1950s [64]. Since then, significant improvements have been made in model resolution, parameterizations, and data assimilation. Institutions such as the European Centre for Medium-Range Weather Forecasts (ECMWF) provide advanced reanalysis datasets, like ERA5, which are instrumental in understanding and modeling atmospheric behavior [28, 51].

Despite these advances, NWP systems face challenges in mountainous regions, where complex terrain leads to localized phenomena that are difficult to model [11]. A study conducted in Switzerland have shown that NWP models struggle to accurately predict Foehn events, a type of wind gust common in Alpine regions, despite extensive research in the underlying physical processes leading to such events remain difficult to predict [3].

In Switzerland, The Swiss Federal Office of Meteorology and Climatology leverages multiple NWP models contribute to ensemble forecasting, supported by high-performance computing infrastructure, including in the Swiss National Supercomputing Centre (CSCS). The recently developed *ALPS* supercomputer is dedicated to advancing atmospheric science through high-resolution simulation capabilities [1, 14]. National efforts are complemented by collaborations with international centers like ECMWF [48]. Even so, high-resolution forecasts remain expensive and not always available to small stakeholders.

Neural Networks for Weather Forecasting. Deep-learning models have recently gained traction as weather prediction tools. Their strength is due to such models tackling the problem from the opposite end. Instead of needing explicit physical modeling, they learn and model

non-linear relationships directly from data. This enables faster models, with lower cost and in many settings, yields equal or better accuracy [2, 17, 33].

Several neural network architectures illustrate the growing potential of deep learning in meteorological forecasting. One prominent example is Gencast, a diffusion-based model trained on the ERA5 reanalysis dataset. This model stands out for its ability to deliver medium-range forecasts up to 15 days ahead, with a runtime of ten minutes. The low-cost and high efficiency makes it an attractive alternative to current methods, in particular for applications requiring constant updates [60]. Another development involves the use of Back-Propagation Neural Networks (BPNNs) for temperature prediction [5]. A study has demonstrated a capacity to significantly reduce forecast error. Likewise, Pangu-Weather, a transformer-based model trained using data from the ECMWF, has outperformed several state-of-the-art NWP systems in benchmark tasks [8]. The field has also witnessed advancements through the application of generative models. Deep generative diffusion models, for instance, have shown improved predictive performance relative to established NWP frameworks [40]. Generative adversarial networks (GANs) trained on radar-based rainfall data in the United Kingdom have shown promising results in enhancing short-term precipitation forecasts. These GAN-based approaches produce high-resolution, realistic simulations of precipitation events, often exceeding the spatial accuracy of traditional radar extrapolation methods [61].

Neural Weather Forecasting in Switzerland. In Switzerland, machine learning has been applied to forecast various weather phenomena. Ranging from Foehn winds in Switzerland [3], to the classification and subsequent prediction of thunderstorms, leading to an improvement of probability detection and a lowering of false alarm rates [58]. Another research explored the combination of LIDAR measurements with weather station data to evaluate wind energy potential in the Swiss Alps [36].

A recent study by the Expertise Center for Climate Extremes (ECCE) at the University of Lausanne focuses on forecasting extreme convective winds, with a particular emphasis on predicting hourly wind gusts up to three days in advance [39]. The workflow is built upon the Neural Weather Model Pangu-Weather, which generates hourly forecasts on a 0.25° global grid. To refine these forecasts at specific locations, ground-based Automatic Weather Stations (AWS) across Switzerland are grouped into five clusters based on inter-station correlation. To generate probabilistic forecasts, the model output is post-processed using a hierarchy of statistical and deep learning techniques. In particular, generalized extreme-value (GEV) distributions are fitted within each AWS cluster to constrain the predictions and ensure statistical robustness. Additionally, a convolutional neural network is used to post-process spatial atmospheric patterns, yielding more accurate forecasts than direct prediction methods. The study concludes that Neural Weather Models, when combined with appropriate post-processing, provide significant added value for the prediction of extreme wind events. This improvement has important implications for the development of early-warning systems, particularly in regions prone to severe thunderstorms.

Forecasting constraints in Mountainous Regions. While some deep learning approaches to precipitation forecasting depend on radar data, mountainous areas such as the canton of Valais face issues due to radar signals often suffering from blockage and attenuation caused by the complex terrain, in turn leading to the underestimation of lower-elevation rainfall [69].

While combining data from multiple sources (e.g., radar, satellite, and surface stations)

can improve forecast reliability [24], such systems may be inaccessible to small-scale users, including local municipalities, farmers, and outdoor event planners, who often lack the infrastructure or funding required.

The mountainous terrain of Valais further complicates precipitation forecasting by creating localized precipitation patterns. This complexity increases the demand for a dense network of ground-based weather stations to ensure spatial coverage and resolution. However, deploying high-end instruments at scale can be costly, reinforcing the need for more affordable solutions.

This has highlighted the importance of low-cost, localized forecasting solutions based solely on time series data from automatic weather stations [37]. The current study contributes to that objective by developing a lightweight forecasting system tailored to the Valais region, using only data collected from automated stations.

1.2. Neural Networks for Time-Series Forecasting. Time series modeling is used in a variety of fields including climate modeling to create forecasts using past data. traditional approaches such as Auto-Regressive methods where future data points are assumed to be a combination of past data are used to produce accurate forecasts [10]. In recent years, Machine Learning methods have been proposed to learn patterns in the data and to produce forecasts. At its simplest, a deep learning method for time-series forecasting is simply that the forecast for one step is the function $f(\cdot)$ using past data for a certain look-back window, with exogenous data for that same window and static exogenous data, meaning data that does not evolve through time, in the case of rainfall prediction, rainfall can be predicted using past rainfall values, various exogenous values such as temperature, pressure or wind and static values such as the coordinates of the weather station [41].

Multi-Layer Perceptrons(MLPs). MLPs have shown potential for time-series forecasting due to their ability to approximate non-linear functions quite well and require only modest computational resources, making them appealing in localized forecasts that require continual retraining [12]. While MLPs do not always outperform traditional statistical methods such as Exponential Smoothing (ETS) or Auto Regressive Moving Average (ARMA), they offer complementary strengths. [29] Specifically, MLPs can model non-linear interactions between variables and are capable of learning seasonal or trend-like behavior given sufficient training.

Hybrid methods. Hybrid approaches combining machine learning models with statistical techniques have demonstrated further improvements. One such example is the fusion of ETS with neural networks, where the statistical model accounts for seasonality and trend, while the neural network learns the residual [57]. These hybrid models could offer a practical path toward interpretable and effective forecasting systems in meteorology. This hybrid method was notably used to forecast wind speed values in Brazil. The results show a better forecast than linear or AI models [16].

Applications to Weather forecasting. In the field of weather prediction, many architectures may be used, one of the most widely used is Recurrent Neural Networks(RNNs) due to their feedback loop keeping dependencies on the long-term. Other architectures are proposed such as Long-Short-Term-Memory(LSTMs) and the even more precise Transductive LSTMs which when applied to weather prediction in Belgium and the Netherlands yield satisfying results [32]. A comparative study of several models' ability to predict rainfall time series in the various cities in the United Kingdom from 2000 to 2020 show that LSTMs yield good

results [6].

1.3. Kolmogorov–Arnold Networks (KANs). Neural networks have emerged as powerful tools for weather forecasting due to their ability to model non-linear interactions from historical data. Their adaptability makes them also especially valuable in the context of climate change, where traditional methods may struggle to capture evolving patterns. These neural networks models not only produce fast and accurate forecasts but also aid in the early detection of extreme weather events, providing critical lead time for response and helping our understanding of climate dynamics [43]. Recent advances such as Gencast, improving early warnings for extreme weather events such as cyclones, enabling local authorities better preparation [60]. In this landscape, Kolmogorov–Arnold Networks (KANs) present a interesting alternative: they retain the forecasting power of deep learning while offering a more lightweight and interpretable structure. Given their performance in other fields, the application to meteorological data and exploring their potential for weather forecasting is relevant.

Kolmogorov–Arnold representation theorem. The Kolmogorov–Arnold representation theorem states that any continuous function can be written as a finite sum of univariate inner and outer functions. The theorem, proposed in 1957 by Soviet mathematician Andrey Kolmogorov lays the groundwork for an approximator built entirely from one-dimensional components [34]. This results implies that every function can be written with univariate functions and sum.

The applications in machine learning are apparent, since any multivariate smooth function boils down to multiple univariate functions. Several studies have tried bringing the Kolmogorov–Arnold representation theorem to neural networks [35, 68]. In 2021, a study demonstrated that the Kolmogorov Superposition Theorem could break the curse of dimensionality [38].

Kolmogorov–Arnold Networks. This theorem was subsequently applied to neural networks, Kolmogorov–Arnold Networks(KANs) are similar to MLPs with a few key differences. While MLPs rely on fixed, non-linear activation functions (such as ReLU or sigmoid) placed at nodes and modulated by linear weights on the edges, in KANs nodes perform simple summations, and the non-linear transformations are instead applied along the edges through learnable, univariate B-spline functions which provides better performance and higher interpretability [45, 67].

As a result, KANs offer a promising alternative to traditional deep networks, especially in tasks where interpretability, smoothness, and structural decomposition are crucial. Their relevance to this study lies in these properties, which align with the goal of building compact, interpretable models for time series forecasting in constrained environments.

Variants. The past year has seen a flurry of new variants, each designed with specific applications in mind. While an exhaustive list would be impractical, this paragraph focuses on architectures most relevant to the present study. For a broader perspective, relevant literature offers an overview of noteworthy architectures that leverage KANs [67].

- **DropKAN** masks a subset of activation-functions, preventing overfitting and improving generalization, for an overall better performance across several UCI machine learning datasets [4].
- **Residual KANs (RKAN)** implement a strategy similar to ResNets (Residual Neural Networks) [27], with most uses geared towards computer vision tasks, The goal is to

supplement Convolutional Neural Networks(CNNs) with RKAN blocks used to help the network extract features while remaining efficient. Offering improvements over several benchmarking datasets such as CIFAR-100, Food-101, Tiny ImageNet [75].

- **Temporal KANs (TKAN)** combine Long Short-Term Memory networks (LSTMs) and KANs, with layers composed of Recurrent KANs, which enables multi-step time-serie forecasting. Tested on hourly traded volumes of Bitcoin, the architecture performed well compared to classic Gated recurrent units (GRUs) or Long Short-Term Memory(LSTMs) [23].
- **Convolutional Kolmogorov-Arnold Networks** are similar to Convolutional Neural Networks (CNNs), which are often used in the field of image recognition. Convolution involves sliding a learnable kernel across input data to extract features by computing dot products at several locations. Convolutional KANs use kernels composed of learnable non-linear functions, which enables more flexibility. This architecture applied an image classification task with the benchmarking Fashion-MNIST dataset yields positive results [9].
- **Multi-layer Mixture-of-KANs (MMK)** is a variant of KAN which uses a mixture-of-experts structure [47] where each input is routed to the most relevant sub-network. In the specific case of MMK, a Mixture-of-KAN (MoK) layer, is created with a gating network. This is applied to time series forecasting where each expert can learn parts of the temporal features for better forecasting. This architecture, when tested on several datasets such as ETT, ECL, Traffic and Weather yields outstanding results [26].

Practical Implementations. Following the initial implementation of the Kolmogorov-Arnold representation theorem in neural networks, the synergy between KANs and scientific research has gained increasing attention. This is evident in the development of dedicated Python packages that facilitate experimentation and integration into applied workflows [44]. In the field of time series forecasting, KANs have been used to predict the volatility index, achieving promising levels of accuracy and interpretability [13]. Similarly, in traffic management, they have demonstrated superior performance compared to traditional MLPs [72]. The release of many python packages, along with a growing ecosystem of tutorials and curated resources [42], is contributing to broader adoption. As a lightweight yet powerful alternative to deep neural networks, KANs present themselves as a promising candidate for meteorological forecasting tasks.

1.4. Spatial Interpolation Techniques. In many environmental and socioeconomic studies, phenomena such as elevation, temperature, or population density are observed at irregular spatial locations due to constraints in data collection methods. However, analysis and modeling often require a complete spatial representation, typically on a regular grid. Since direct measurements at every grid point are rarely feasible, spatial interpolation becomes essential to estimate values in unmeasured areas. This step enables the transformation of discrete, scattered data into continuous surfaces suitable for visualization and further analysis in GIS environments [50].

Rainfall Interpolation. Even products derived from remote sensing, such as weather radar, still rely on interpolation and modeling to refine their estimates. This is often the case for radar-based precipitation products which often tend to underestimate rainfall in complex

terrain like the Alps. To improve rainfall estimates, several correction strategies are applied. Radar bias can be reduced by comparing radar intensities with rain-gauge measurements treated as ground truth. Conversely, gauge observations can be spatially interpolated using radar data as an external covariate in the regression. In both cases, one data source is taken as primary. A genuinely integrated approach allows either source to dominate locally, with the overall aim of minimizing overall uncertainty rather than simply correcting one dataset with the other [55]. In Switzerland, this integration is realized in the advanced radar product *CombiPrecip*. Five-minute radar scans are first aggregated to hourly sums, a similar workflow is applied to gauge readings. The resulting field is obtained through *co-kriging with external drift*, a geostatistical technique that interpolates rain gauge values with radar as the external drift. In regions where gauge support is sparse—for instance outside Swiss borders—the algorithm assigns greater weight to radar observations, likewise, kriging is only performed outside of Swiss borders if rainfall was detected by radar. This step is crucial as Kriging tends to have a smoothing effect with rainfall decreasing over distance when no other information is specified. Leave-One-Out Cross-Validation (LOOCV) is then used to quantify local interpolation errors by omitting each gauge in turn, interpolating its value, and comparing the result with the true measurement. The precipitation field produced by this workflow has proved robust and underpins a range of applications, including rainfall nowcasting [66].

Kriging. This widely used geostatistical interpolation method leverages the concept of spatial autocorrelation, assuming that observations closer together are more alike than those further apart. By modeling the spatial structure of the data through a variogram, Kriging provides not only predictions at unsampled locations but also estimates of the associated uncertainty [56]. A notable extension, Kriging with External Drift (KED), incorporates auxiliary variables to guide the interpolation process [25]. For example, incorporating a digital elevation model (DEM) as an external drift can significantly improve predictions, particularly when the phenomenon of interest is strongly linked to topographic features. In the case of the Valais region, it has been shown that using elevation as a covariate within KED enhances the accuracy of rainfall estimates, outperforming other interpolation approaches [20]. Open-source libraries such as **GSTools** [53] simplify the implementation of Kriging and provide additional tools for variogram estimation, model selection, and simulation of spatial fields. [53] These tools can be integrated into forecasting pipelines to support spatial generalization from localized station data.

2. Theoretical foundation.

2.1. Kolmogorov-Arnold Networks. The Kolmogorov–Arnold theorem states that any continuous multivariate function on a bounded domain can be expressed as a finite composition of continuous one-dimensional functions. This theorem is existential and non-constructive, it guarantees that such functions exist but does not provide a method to construct them.

Formally,

$$(2.1) \quad f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \phi_{p,q}(x_p) \right),$$

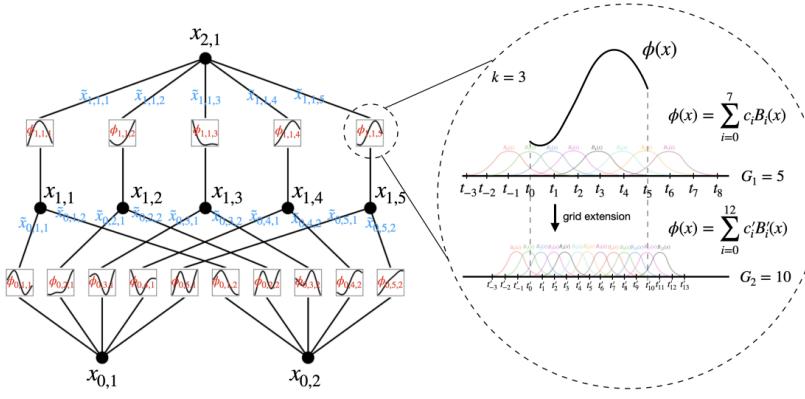


Figure 1: Schematic representation of a Kolmogorov–Arnold Network (KAN). **Left:** The network structure shows activation functions applied on edges instead of nodes. In this example, there are $n = 2$ input nodes and $2n+1 = 5$ hidden layer nodes. **Right:** Each edge’s activation function is parameterized as a B-spline. G denotes the number of intervals, with each spline fit over the aforementioned intervals. Source: Liu et al. [45].

where $\phi_{p,q} : [0, 1] \rightarrow \mathbb{R}$ and $\Phi_q : \mathbb{R} \rightarrow \mathbb{R}$ are continuous functions. The index $p = 1, \dots, n$ refers to the n *input* variables, while $q = 1, \dots, 2n + 1$ labels the *output* functions summed together to form an approximation.

While this theorem did not immediately lend itself to machine learning applications, it laid the conceptual foundation for Kolmogorov–Arnold Networks (KANs), a recently proposed neural architecture.

KANs share a similar structure with Multi-Layer Perceptrons (MLPs), typically modeled as layered graphs. Information flows from an input layer, through one or more hidden layers, and finally into an output layer. As shown in Fig. 1, a KAN layer from the bottom to the top includes n input nodes, $2n + 1$ hidden nodes, and an output layer.

Whereas classic MLPs use weights on edges and apply activation functions at the nodes, KANs place activation functions directly on the edges. That is, each edge from input node p to hidden node q carries its own activation function, denoted $\phi_{p,q}$. This setup is clearly illustrated in Fig. 1, where each connection between two nodes has its own function. Each activation function $\phi_{p,q}$ is parameterized as a combination of B-splines.¹

Each activation function can be formally expressed as

$$\phi_{p,q}(x) = \sum_{k=1}^K w_{p,q,k} B_k(x),$$

where $B_k(x)$ is a B-spline function and $w_{p,q,k}$ is a coefficient to be trained when fitting a

¹A *B-spline* (short for basis spline) is a polynomial function commonly used for approximating other functions over an interval. In the case of cubic B-splines, each segment is a third-degree polynomial that is smoothly joined with its neighbors. Fig. 1 (Right) illustrates several spline functions over a domain divided into G intervals. These basis functions are combined to construct smooth approximations of more complex functions [15].

model. The number K corresponds to the total number of spline basis functions used. Typically, degree-3 (cubic) B-splines are chosen, offering a good balance between flexibility and smoothness, with cubic functions being continuous up to their second derivative.

An entire KAN layer can be described as a matrix $\Phi = [\phi_{p,q}]$, where each function $\phi_{p,q}$ represents a univariate function connecting an input node p to an output node q . Nodes are fully connected² between layers via an edge, and each edge carries an activation function $\phi_{p,q}(x)$. Because KANs optimize a matrix of one-dimensional functions rather than a single high-dimensional one, training tends to be more efficient than traditional MLPs while having comparable performance.

A full KAN with L ³ layers is expressed as:

$$y = \text{KAN}(x) = (\Phi_L \circ \Phi_{L-1} \circ \dots \circ \Phi_1)(x),$$

where Φ_L produces the final output, and the composition operator \circ denotes that the output of each layer feeds into the next. This proposed architecture [45], offers an alternative to traditional neural networks.

2.2. Kriging Interpolation.

Ordinary Kriging (OK). In a spatial field where data points are irregularly distributed, each observation is associated with a specific location denoted as \mathbf{s}_i , typically expressed in latitude and longitude coordinates. Ordinary Kriging is a method used to estimate the value at an unknown location \mathbf{s}_0 by computing a weighted linear combination of the n known observations. Formally, this is expressed as:

$$(2.2) \quad \hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i), \quad \sum_{i=1}^n \lambda_i = 1,$$

where $Z(\mathbf{s}_i)$ represents the measured value at location \mathbf{s}_i , the weight λ_i is assigned to each observation, and $\hat{Z}(\mathbf{s}_0)$ is the predicted value at the unsampled location. The constraint that the sum of weights must be equal to one ensures that if all observations are equal to a constant, the estimate will also return that same constant. Weights may in certain cases be negative.

What distinguishes Kriging from other methods such as Inverse Distance Weighting (IDW) is that the weights are not simply based on distance, but also incorporate the spatial structure of the observed field. Spatial structure shows how the similarity between observations evolves with distance. For example, consider a location s_i and two distances h_1 and h_2 . If the variable of interest changes by a certain amount over distance h_1 , and changes even more over h_2 , this indicates that similarity decreases as distance increases. The underlying assumption is that locations closer together tend to have more similar values. The spatial structure quantifies similarity and describes how quickly it declines over distance. This relationship is captured by the *semi-variogram*:

²A fully connected layer means that every node in one layer is connected by an edge to every node in the next layer.

³The number of layers L is a tunable hyperparameter that can be adjusted.

$$(2.3) \quad \gamma(h) \equiv \gamma(|s_i - s_j|) = \frac{1}{2} \text{Var}[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)],$$

which quantifies the semi-variance between pairs of points \mathbf{s}_i and \mathbf{s}_j separated by a distance h . A model is then fit to the variogram, akin to fitting a regression line, but using parametric forms such as the Spherical, Exponential, or Gaussian models. These models describe how the semi-variance increases with distance and eventually levels off, reflecting how points too far apart are no longer dependent. Each form has a different shape: for instance, the Spherical model levels off more rapidly, while the Exponential model approaches its maximum more gradually. Choosing the right model helps accurately represent the spatial behavior of the data.

The optimal weights λ_i^* are obtained by solving a system of equations minimizing the Kriging variance. This system, using a Lagrange multiplier μ , is defined as:

$$(2.4) \quad \begin{aligned} \sum_{j=1}^n \lambda_j \times (|s_i - s_j|) + \mu &= \gamma(|s_i - s_0|), \quad i = 1, \dots, n, \\ \sum_{j=1}^n \lambda_j &= 1 \end{aligned}$$

Solving the equation system (2.4) yields the optimal weights. Making it simple to predict unknown values at new locations using nearby values.

Figure 2 shows an example of a semi-variogram fit to precipitation data. Despite involving 48 stations, there are not 1128 pairs due to the use of binning. During the fitting process, distance intervals (or bins) are defined, and the average semi-variance inside each bin is computed. The variogram model is then fit to these bin averages, which helps reduce noise and stabilize the curve.

Kriging with External Drift (KED). Ordinary Kriging fits a model using only the spatial arrangement of data points and a variogram. In comparison, Kriging with External Drift (KED) assumes that the variable of interest varies not only with distance but also in relation to an auxiliary variable $q(s)$, such as elevation when modeling precipitation.

The observation $Z(s)$ at a location s can be rewritten as:

$$(2.5) \quad Z(\mathbf{s}) = \beta_0 + \beta_1 q(s) + \varepsilon(s),$$

where β_0, β_1 are unknown coefficients, $q(s)$ is the value of an auxiliary variable at a location s . The term $\beta_0 + \beta_1 q(s)$ is known as the *drift*. Typically written as $\mu(s)$, it represents the trend driven by the auxiliary variable. The residual $\varepsilon(s)$ is assumed to have a mean of zero, to be isotropic (not depending on direction), and to depend only on the distance between points.

To enforce consistency with the external drift, a second constraint is introduced along

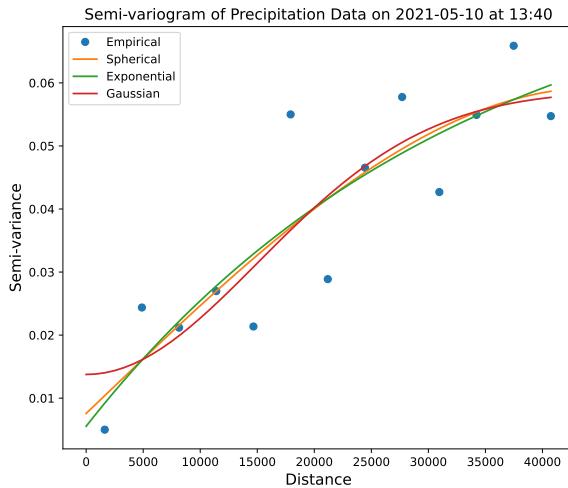


Figure 2: Semi-variogram of precipitation data on May 10, 2021 at 13:40. Several parametric models are shown: Spherical, Exponential, and Gaussian.

with an additional Lagrange multiplier μ_1 for the drift. The *KED system* becomes:

$$(2.6) \quad \begin{aligned} \sum_{j=1}^n \lambda_j \gamma(|s_i - s_j|) + \mu_0 + \mu_1 q(s_i) &= \gamma(|s_i - s_0|), \quad i = 1, \dots, n, \\ \sum_{j=1}^n \lambda_j &= 1, \\ \sum_{j=1}^n \lambda_j q(s_j) &= q(s_0) \end{aligned}$$

These constraints ensure that the interpolated trend matches the auxiliary variable at the prediction location. Solving system (2.6) yields optimal weights λ_i^* , which account for both the nearby observations and the spatial pattern of the auxiliary variable $q(s)$. Empirical study in the Canton of Valais by Tobin et al. [69] has shown that incorporating a digital elevation model as an external drift improves the accuracy of precipitation interpolation compared to other interpolation methods.

3. Dataset and Exploratory Data Analysis. This chapter provides an overview of the data used and the initial exploratory analysis performed to understand its structure and characteristics. In Section 3.1, we describe the dataset, which consists of time series data collected from automatic weather stations located in the Valais region. Section 3.2 presents the exploratory data analysis (EDA), where we examine the descriptive statistics and investigate the behavior of our variables. Section 3.3 focuses on the presence and patterns of missing values.

3.1. Data Description. The dataset utilized in this study was obtained from MeteoSwiss, the Swiss Federal Office of Meteorology and Climatology [49]. It comprises meteorological data recorded by automatic weather stations distributed throughout the Canton of Valais, Switzerland. The observation period spans from January 1, 2019, to December 31, 2023.

The network of Automatic Weather Stations (AWS) in Switzerland is known as *SwissMetNet*. The stations monitor weather conditions across the country and transmit their measurements to a centralized MeteoSwiss database, with each station recording a range of variables at a temporal resolution of 10 minutes. Within the Canton of Valais, a total of 54 stations are deployed. Among these, 12 record all weather variables listed in Table 1, while 2 record all except atmospheric pressure, 1 omits wind measurements, and the remaining 39 stations are dedicated exclusively to precipitation. The high number of stations in Valais, especially those focused on precipitation, highlights the region's complex topography. The canton of Valais is also particularly vulnerable to extreme weather events such as floods, especially in the Rhone River, where floods can heavily damage infrastructure [22]. Therefore, a denser network of stations is required to ensure reliable data.

In the data aggregation process used to build a comprehensive data warehouse, MeteoSwiss collaborates with various partners, including private entities, cantonal agencies, and even international organizations. These partners contribute to broader data coverage. However, in the present study, only SwissMetNet stations are used.

All instruments in the SwissMetNet network comply with World Meteorological Organization (WMO) standards [74]. These standards ensure consistency, precision, and comparability across time and space. In Switzerland, standards are adapted to reflect particular environmental conditions encountered [19]. These certified methods provide confidence in the quality of the dataset. Although some missing values or measurement biases may still occur, the consistent methodology applied across stations ensures reliable results.

The data used in this study comes from two sources, a time series dataset containing meteorological observations recorded every 10 minutes and a dataset containing static metadata for each weather station. From the static dataset, only the location and altitude are used in the modeling. The full list of available static metadata is provided in Appendix A.

Variables used in the model are detailed below. Table 1 also provides a summary of each variable, including a short description and a range of possible values.

Precipitation. Measured in millimeters (mm), the total amount of precipitation received over a 10-minute interval is recorded. The value represents the height of water over one square meter, meaning that 1 liter of water corresponds to 1 millimeter of rain.

Precipitation is recorded using two types of instruments: tipping-bucket rain gauges and weighing gauges. The tipping-bucket model counts how often a small container fills and tips, while the weighing gauge measures the total weight of collected water, offering higher accuracy. Precipitation sensors are also heated, ensuring that snow and hail are correctly measured.

Regarding uncertainty, several factors—such as wind and evaporation—can impact accuracy. Estimates suggest this error ranges from 4% to 35%, depending on the station and the season [71]. Overall, rain gauge measurements are considered robust and reliable.

Temperature. Temperature is measured in degrees Celsius ($^{\circ}\text{C}$) and recorded every ten minutes as an instantaneous value. Several types of temperature readings exist. Air temperature is measured two meters above the ground. Surface temperature is measured five

Table 1: Variables used in the model, including their symbolic representation, description, and expected value types or ranges.

Variable	Symbol	Description	Values
<i>Time</i>	t	Timestamp in format: YYYYMMDDHHMM	—
<i>Altitude</i>	z	Station altitude above sea level	Positive integer (m)
<i>Easting</i>	x	Easting coordinate in the LV95 system	Positive integer (m)
<i>Northing</i>	y	Northing coordinate in the LV95 system	Positive integer (m)
<i>Precipitation</i>	P	Total precipitation over 10-minute intervals	Float ≥ 0 , 1 decimal (mm)
<i>Humidity</i>	HM	Relative humidity at 2 m above ground	Float in [0, 100], 1 decimal (%)
<i>Atmospheric Pressure</i>	AP	Sea-level corrected atmospheric pressure	Float ≥ 0 , 1 decimal (hPa)
<i>Temperature</i>	TM	Air temperature at 2 m above ground	Float ($^{\circ}$ C), 1 decimal
<i>Wind North</i>	V	Vertical Wind vector component	Float (m/s), 1 decimal
<i>Wind East</i>	U	Horizontal Wind vector component	Float (m/s), 1 decimal
<i>Speed</i>	s	Computed magnitude of the wind vector from U and V	Float (m/s), derived
<i>Direction</i>	θ	Computed direction of the wind vector in degrees from north	Float (degrees), derived

centimeters above the ground, and ground temperature is recorded at various depths, typically 5, 10, and 20 centimeters below the surface.

In this study, air temperature is selected, as it highly influences precipitation, with warmer air being able to hold more moisture [54]. The sensors used are mainly of two types: Thygian and Rotronic. Both are designed to operate under challenging weather conditions, with ventilation and shielding systems to minimize bias due to radiation or wind.

Air Pressure. Air pressure is recorded in hectopascals (hPa) as an instantaneous value of the weight of the air column above a given location. Since there is less air above at higher altitudes, station elevation significantly influences measurements. However, pressure is also affected by weather conditions. To accurately compare pressure between stations, the influence of altitude must be corrected through a process known as *reduction*, which adjusts the measured pressure to its equivalent at sea level. This adjustment is complex because air is compressible and the relationship between altitude and pressure is non-linear. MeteoSwiss provides raw and reduced pressure values as part of their datasets. Two main reduction methods are applied depending on station elevation. For stations located below 600 meters, the more accurate method, known as *QFF*, is applied. It incorporates current weather

conditions using temperature and humidity to reflect more closely the relationship between altitude and pressure. For stations above 600 meters, the *QNH* method is used. It is based on the International Standard Atmosphere (ISA) and applies a fixed reduction depending only on altitude. While QNH may be slightly less precise than temperature-adjusted reductions, the difference is often small, and QNH remains well suited for weather prediction [46].

In this study, for consistency across stations, the simpler QNH method was used, even for stations located below 600 meters.⁴

Relative Humidity. In this study, humidity refers to relative humidity, expressed as a percentage (%) of water vapor contained in the air. Two types of humidity measurements exist: absolute humidity, which refers to the actual amount of moisture in the air, and relative humidity, which compares current moisture to the maximum possible at a given temperature. Relative humidity is correlated with temperature as warmer air can hold more moisture. The sensors used for humidity measurement are the same as those used for temperature.

Wind Speed and Direction. Wind data is recorded every ten minutes and includes both the average wind speed and direction. It is measured using either spinning cup anemometers with wind vanes or ultrasonic sensors, which detect wind based on differences in signal travel times between sensors. Wind speed is expressed in meters per second (m/s), and wind direction in degrees ($^{\circ}$). Note that in meteorology, wind direction refers to the direction *from which* the wind is blowing. For instance, a westerly wind (blowing from west to east) has a direction of 270° .⁵

Although wind direction in degrees is convenient for visualization, it can be problematic in numerical models.⁶ To address this, wind is expressed as vector with the following components:

$$(3.1) \quad \begin{aligned} U &= -(s \times \sin(\theta_{rad})), \\ V &= -(s \times \cos(\theta_{rad})). \end{aligned}$$

Here, U and V represent the eastward and northward components, respectively, s is the speed in m/s, and θ is the direction converted to radians.

Location. Location is expressed using the Swiss Coordinate System, known as LV95⁷. Developed specifically for Switzerland, this system has its origin in Bern, with coordinates $E = 2,600,000$; $N = 1,200,000$, ensuring all values within Switzerland are positive.

Coordinates are expressed as Easting and Northing in meters, unlike the global WGS84 system, which uses latitude and longitude. For example, the Sion weather station is located at $E = 2,591,633$; $N = 1,118,584$.⁸

⁴Only two stations in the entire Canton of Valais measure pressure and are located below 600 meters, Sion and Evionnaz. These are therefore the only stations for which QFF values are available. Since all stations provide QNH values, using it ensures consistency across the dataset.

⁵Wind direction is calculated clockwise from the top of the vertical axis (i.e. from the north). Thus, a wind blowing from west to east will have an angle of 270° from the vertical axis, while a 0° wind will correspond to a northerly wind (from north to south).

⁶Indeed, directions of 359° and 0° are numerically far apart, but in reality, they are nearly the same.

⁷LV95 stands for *Landesvermessung 1995*, which means “land survey 1995”. It refers to the updated Swiss national coordinate system established after the 1995 geodetic survey.

⁸For more details on the LV95 coordinate system, see Wiget et al. [73] (in German) and the official Swisstopo

Finally, a Digital Elevation Model (DEM) of Switzerland is incorporated into the analysis. Sourced from the Federal Office of Topography swisstopo [18] and with a spatial resolution of 200 meters, the DEM provides elevation data for regions not directly covered by weather stations. This dataset is used in spatial interpolation tasks, particularly for modeling and visualizing precipitation across complex terrain. Figure 3, generated using the `rayshader` R package [52], shows the topography of Switzerland, highlighting the Alpine region and the challenging nature of the study area.

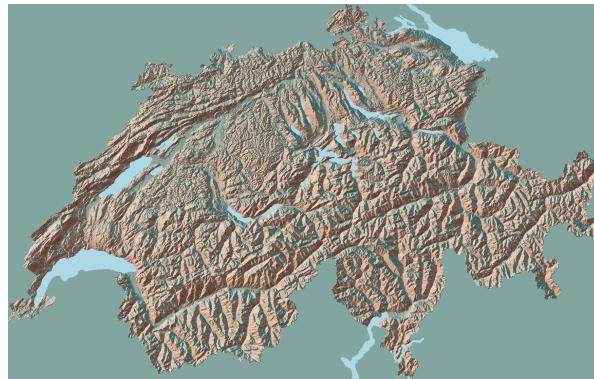


Figure 3: Terrain map of Switzerland with major lakes. Elevation data has a spatial resolution of 200 m × 200 m.

3.2. Exploratory Data Analysis. The Exploratory Data Analysis is structured into several parts: it begins with descriptive statistics 3.2.1, followed by correlation matrix analysis 3.2.2, and visualizations 3.2.3, all aimed at providing a comprehensive overview of the dataset.

3.2.1. Descriptive Statistics. Table 2 presents the descriptive statistics of the key variables. For improved interpretability, wind speed (s) and wind direction (θ) are shown instead of the vector components. However, the original wind vectors are used in the final modeling.

Precipitation. The largest 10-minute rainfall in the record is 44.6 mm at Arolla. For comparison, the official Swiss record according to MeteoSwiss in Lausanne on June 11, 2018 with 41 mm, casting doubt on the accuracy of the value recorded by the Arolla station. The top 3 highest precipitations are all recorded in Arolla, in fourth position, Emosson recorded a total of 22.6 mm on July 21, 2020.

Regarding the longest periods without any rain, Table 13 in the Appendix shows the longest time without any amount of precipitation, with the top value in Ulrichen. registering more than a month without any rain.

documentation available at <https://www.swisstopo.admin.ch/en/the-swiss-coordinates-system>, accessed May 25, 2025.

Table 2: Descriptive Statistics

	<i>P</i>	<i>HM</i>	<i>AP</i>	<i>TM</i>	<i>s</i>	θ	<i>z</i>
mean	0.02	69.00	1018.37	7.96	2.04	144.12	1464.23
std	0.12	20.83	8.24	8.43	2.05	-	498.39
min	0.00	4.40	976.70	-24.70	0.00	-	461.00
25%	0.00	52.80	1014.00	1.60	0.71	-	1204.25
50%	0.00	70.70	1019.30	7.60	1.31	-	1536.50
75%	0.00	87.60	1023.90	13.90	2.57	-	1849.50
max	44.60	100.00	1044.90	38.00	21.09	-	2300.00

The wettest station overall is Clusanfe (1928,m,a.s.l.) with a total 11,353 mm over 5 years, averaging 6.2 mm/day. By contrast, the driest is Mattsand (1230,m) with 2,673 mm, a good example of the dry climate of valleys of inner-Valais⁹, which receive less precipitation due to their protection from humid air masses originating from both the north and the south [77]. Both stations can be seen in Figure 4



Figure 4: Photograph of the stations of Clusanfe (Left) and Mattsand (Right)

Temperature. Temperatures span a 63°C range from -24.7°C (Ulrichen) to 38°C (Sion, July 24, 2019). Sion not only sets the absolute maximum but dominates the 10 minutes temperature averages. In fact, the first 50 hottest temperatures recorded are all in Sion, the next hottest reading is in Martigny with 36.5°C. On the colder end, the second lowest temperature -19.8°C recorded in Turtmann, a station at an altitude of 2,180 m. emphasizing the relationship between temperatures and altitude.

Humidity. While relative humidity can theoretically drop to zero percent, this is practically not feasible. Even in extremely dry regions such as Death Valley, CA, humidity never reaches zero [62]. In our dataset, the lowest observed relative humidity is 4.4%, recorded in Simplon-Dorf.

In contrast, maximum humidity levels are quite common, as air can become fully saturated with water vapor, resulting in fog or cloud formation. For example, the station at Grächen recorded 8,965 ten-minute intervals at 100% humidity, which corresponds to more than 62

⁹For more information, visit: MeteoSwiss, *The Climate of Switzerland*, <https://www.meteoswiss.admin.ch/climate/the-climate-of-switzerland.html>, accessed May 19, 2025.

days. In the Appendix, Table 17 presents by station the total number of observations where humidity reached 100%.

Pressure. Sea-level pressure is stable across stations and over time, averaging 1018.37 hPa with a variance of 67.4 hPa. The overall stability gives us reassurance that any interpolation efforts of pressure will return satisfying results.

Regarding extreme values, the ten lowest pressure occurrences in the dataset are linked to two distinct weather events. The first took place in December 2020, when a deep depression over Scotland caused widespread low-pressure readings across Valais.¹⁰ The second event occurred on October 20, 2023, in Évolène, where the lowest pressure in the dataset (976.7 hPa) was recorded. This was caused by Storm Céline, which brought warm air masses, triggered Foehn winds, and led to heavy precipitation.¹¹ The complete list of the ten lowest pressure values is provided in Appendix Table 16.

Wind. Wind is summarized in Table 2 using speed and direction. Statistics regarding wind direction are limited to the mean, since other descriptive measures (e.g., variance or percentiles) are generally not meaningful for circular data. Even the mean direction can sometimes be misleading.¹² In this dataset, the average wind direction across stations is 144°, indicating an origin from the south-southeast.

Mean 10-minute wind speeds are 2.04 m/s, and the 75th percentile indicates that 75% of observations fall below 2.57 m/s. Yet, the station in Visp recorded an average wind speed of 21 m/s or 75 km/h in 2019. Given that gusts¹³ exceeding 70 km/h are considered a moderate hazard, a sustained average of 75 km/h could pose a serious risk to infrastructure.

Altitude. Altitude values are relatively well distributed across the dataset, with an average station elevation of 1,464 meters, stations are mostly located on higher ground, reflecting the mountainous terrain of the Canton of Valais, this is important to consider during modeling, as weather conditions can substantially vary with elevation.

The lowest station, Martigny (462 m.), is a precipitation-only station situated low in the Rhône Valley. In contrast, the highest station, Bruchji, is located at 2,300 meters near the Aletsch Glacier, offering measurements in an alpine area.

3.2.2. Correlation Matrices. Figure 5 shows the correlations between variables for the city of Sion. Several relationships are apparent. The strongest correlation is between the two wind components (U and V), which is expected. Notably, the correlation is positive, indicating that both components tend to increase together. This suggests that wind often blows along a northeast–southwest axis. In Sion’s case, this is consistent with its geography, the city lies in the Rhône Valley, and bordered by mountains to the northwest and southeast.

Humidity (HM) also shows a moderate positive correlation with wind components. Although this relationship is influenced by multiple factors, prior studies have shown that wind

¹⁰Weather in December 2020 (French only), available at: https://www.meteosuisse.admin.ch/dam/jcr:ebb31915-b85a-41e2-8c7e-badec7b381da/202012_f.pdf, accessed May 21, 2025.

¹¹Mise à jour – Avis précipitations, foehn tempétueux (French only), available at: <https://www.meteosuisse.admin.ch/portrait/meteosuisse-blog/fr/2023/10/mise-a-jour-avis-précipitations-foehn-tempetueux.html>, accessed May 21, 2025.

¹²For example, if winds are evenly split between opposing directions, the resulting average will suggest an absence of wind.

¹³A wind gust is defined as the maximum wind speed measured over a 1-second interval.

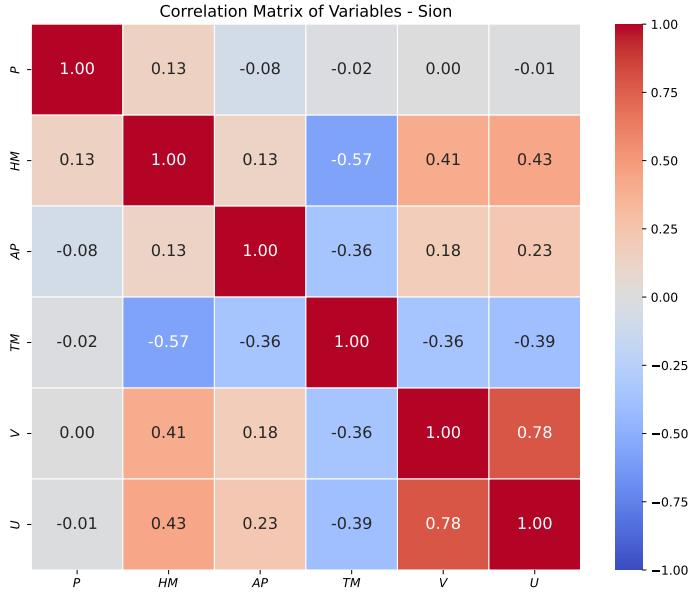


Figure 5: Correlation matrix for Sion station: P (Precipitation), HM (Humidity), AP (Pressure), TM (Temperature), V (North wind), U (East wind).

speed generally has a weak but nevertheless existing relationship with relative humidity [76]. Additionally, humidity is negatively correlated with temperature (TM), which is expected since warmer air can hold more water vapor, lowering relative humidity.

An interesting result is the negative correlation between atmospheric pressure (AP) and temperature. One might expect a positive relationship, as lower pressure is typically associated with storms. However, the opposite appears here, indicating that a local effect may be taking place.

Finally, the relationship between wind and temperature illustrates why correlation analysis should be interpreted with caution. The temperature variable used here is not perceived temperature, but the actual measured air temperature. One may erroneously interpret that a negative relationship implies that wind directly affects temperature. Instead, both variables likely respond to another variable, rather than influencing each other.

Another correlation matrix, Figure 6 helps illustrate precipitation correlations across stations in Valais. In general, stations that are geographically closer seem to display higher correlation values. For example, the station of Evolène (EVO) in the Val d'Hérens and Mottec (MTE) in the neighboring Val d'Anniviers show a higher correlation, which is expected given that the two valleys are separated by only about 10 kilometers.

To assess whether distance influences precipitation correlation, Figure 7 shows a scatter plot comparing for each station, their distance to one another with their corresponding cor-

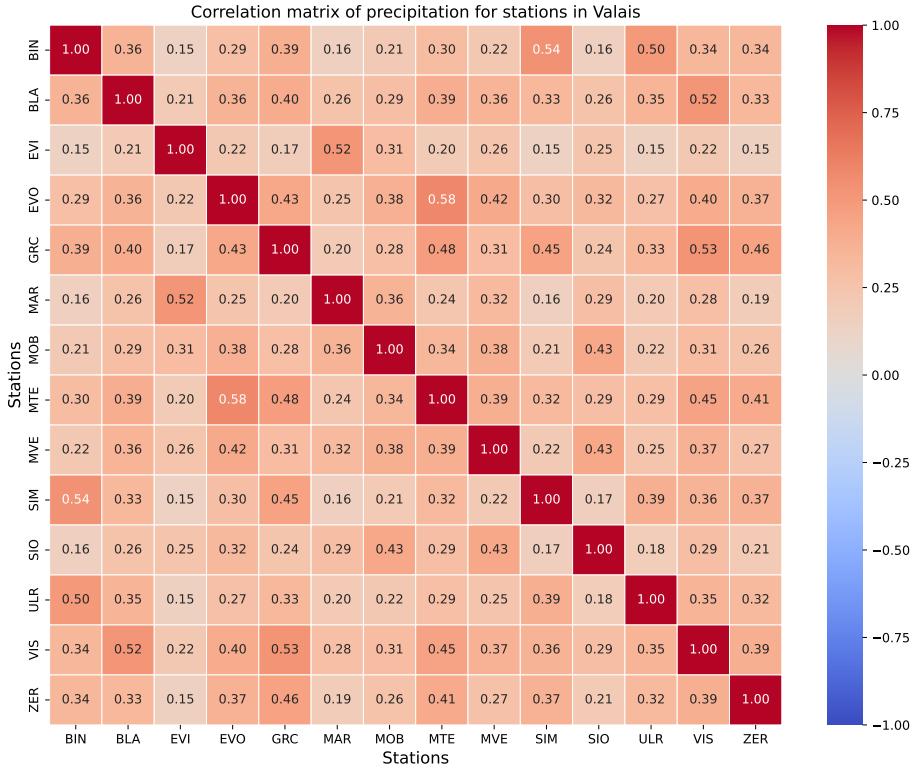


Figure 6: Correlation matrix of precipitation values between 14 selected stations out of 54. Stronger correlations often occur between stations in nearby valleys.

relation coefficients. While some station pairs at shorter distances show high correlation, no clear trend emerges overall. This suggests that precipitations in Valais are similar across the Canton, with local variations not being dependent on distance.

3.2.3. Visualizations. This section explores various visualizations from the dataset, with an emphasis on precipitation. Figure 8 displays all available variables for the station of Sion over time. Red lines indicate monthly averages. As expected, the monthly average precipitation is frequently close to zero, highlighting the highly sparse nature of this variable. This sparsity poses challenges for modeling rainfall events. In contrast, other variables (e.g., temperature, pressure) exhibit clearer seasonality.

Figure 9 focuses on precipitation data from five stations across Valais. Seasonal variation is evident, with certain periods receiving more frequent rainfall and others remaining relatively dry. The selected stations offer a diverse geographic representation. From city centers in the valley such as Sion, Sierre, and Visp, to mountainous areas like Derborence and Zermatt. The

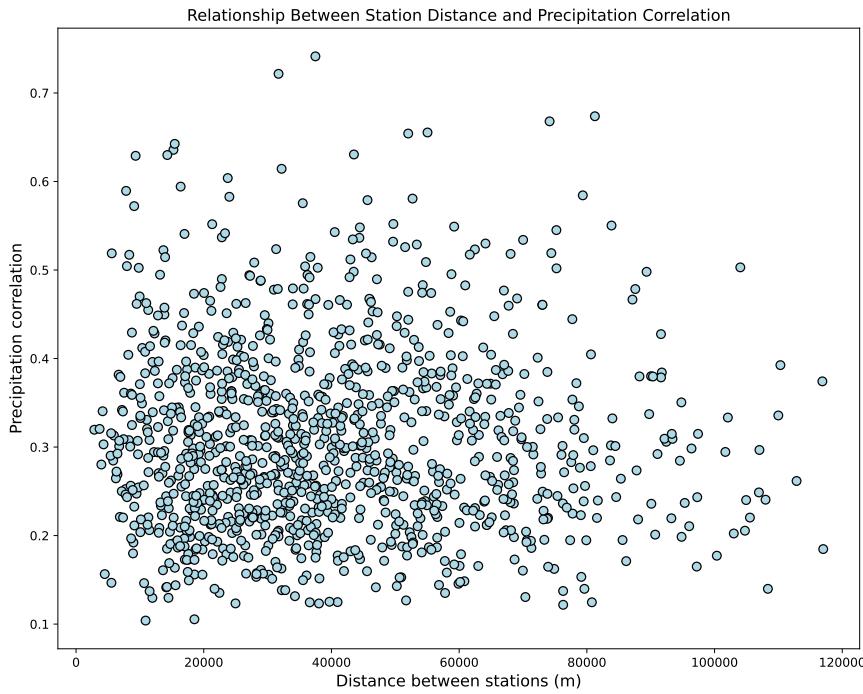


Figure 7: Relationship between the geographic distance and precipitation correlation for station pairs. No strong spatial trend is observed.

station of Derborence, near a mountain lake, exhibits notably high rainfall. Sion, an urban center, shows lower overall precipitation but higher peaks, attributed to summer thunderstorms. This pattern is consistent with findings from a study by the Faculty of Geosciences and Environment at the University of Lausanne, which demonstrated that thunderstorms tend to be more intense over cities [70].

Precipitation data across the dataset is highly sparse. Depending on the station, between 85% and 96% of all 10-minute measurements contain no rainfall. Figure 10 demonstrates how this sparsity can be mitigated using a rolling window. Here, a six-hour rolling sum of precipitation is computed for the station of Sion. This approach smooths the data while keeping rain patterns intact. Indicating that it may be appropriate for modeling purposes.

Figure 11 shows the cumulative sum of precipitation for the same five stations. The cumulative view makes inter-station differences more apparent. Derborence, in particular, receives significantly more rainfall than other locations. Meanwhile, stations in the Rhône Valley (Sion, Sierre, Visp) display similar cumulative totals, suggesting a relatively uniform valley climate. Zermatt, despite its elevation, registers a similar amount of rainfall.

Finally, Figure 12 examines the relationship between station altitude and total precipitation over the study period. Based on previous visualizations, one may expect a clear correlation between the two variables. This is not the case. However, the scatter may hint at the presence of distinct precipitation regimes depending on altitude.

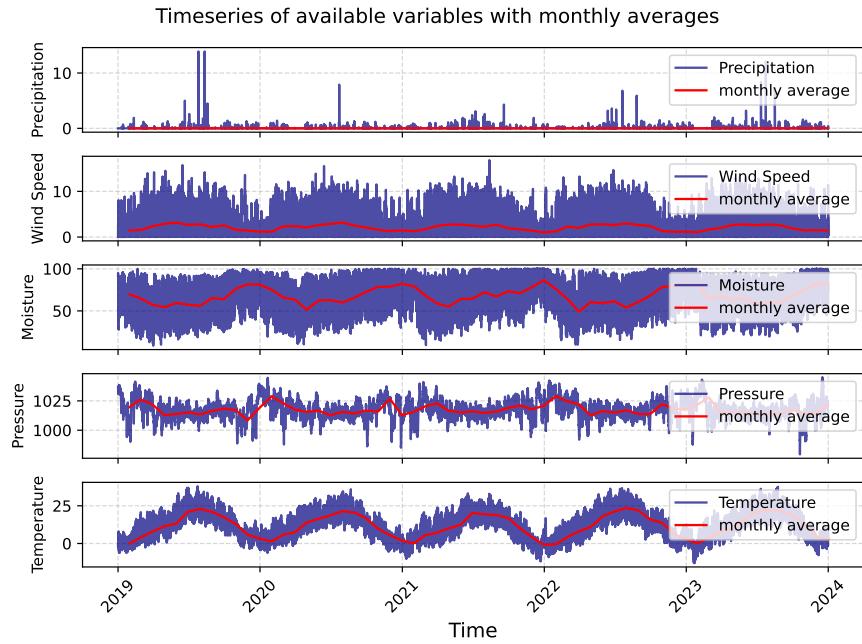


Figure 8: Time series of all variables for Sion station. In red, monthly averages show sparsity in precipitation and seasonal patterns in remaining variables.

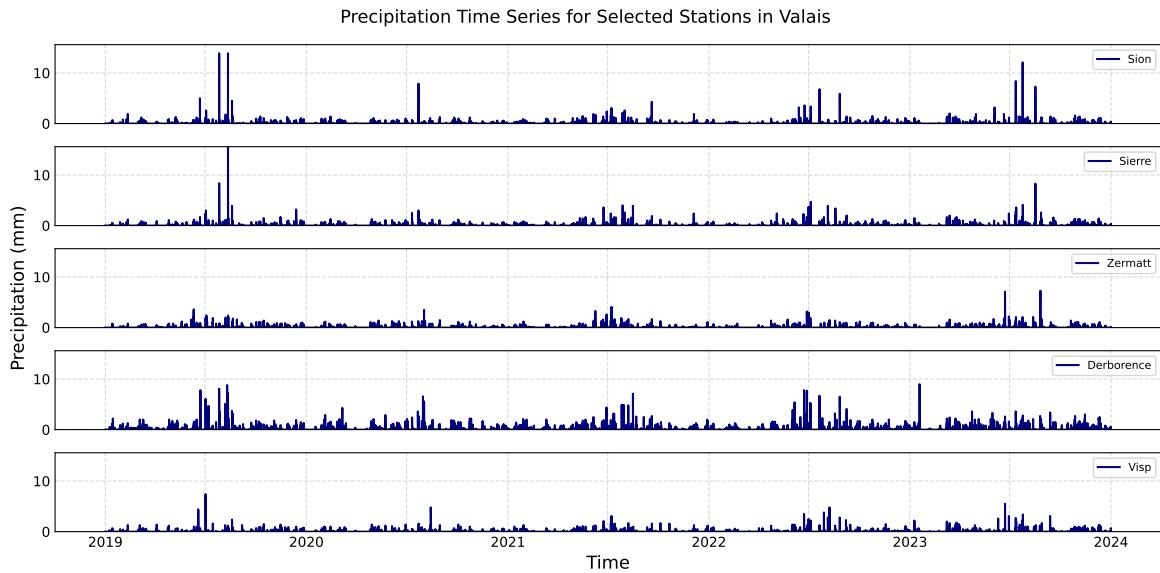


Figure 9: Total precipitation over 10-minute windows for selected stations. The stations span a range of altitudes and locations, from cities to alpine areas.

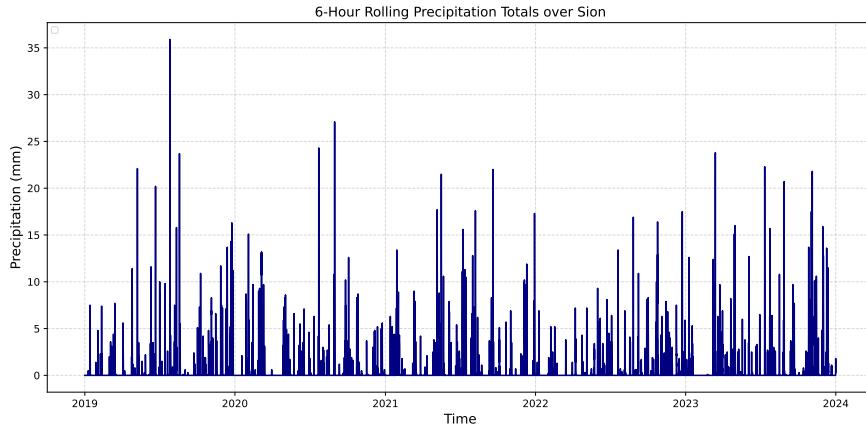


Figure 10: Rolling sum over the past 6 hours precipitation at the Sion station.

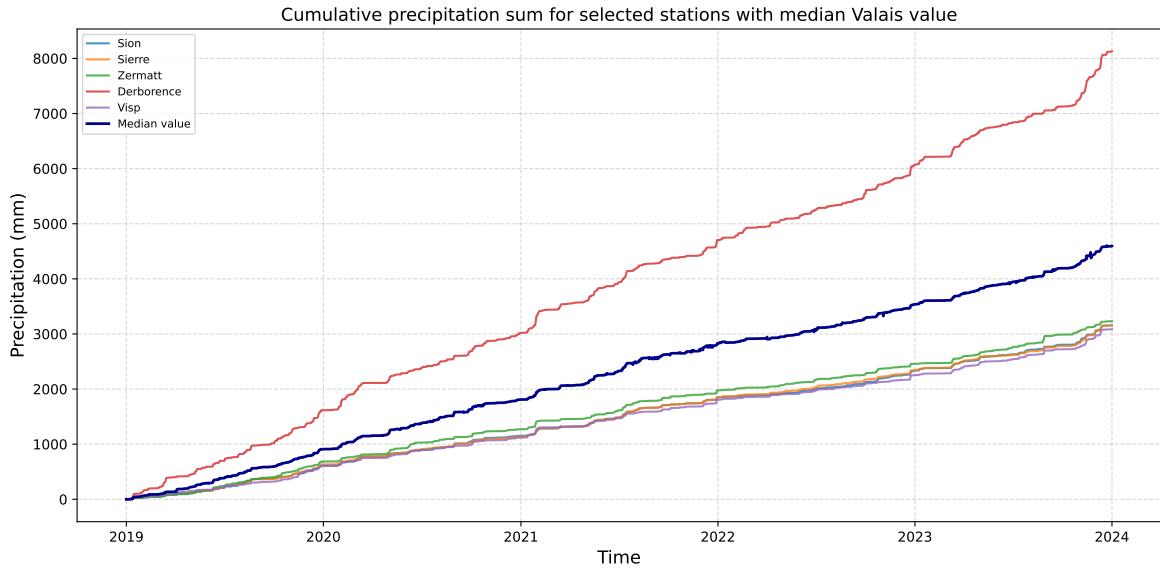


Figure 11: Cumulative sum of precipitation over the studied period.

While total precipitation does not exhibit a clear correlation with altitude, Figure 13 shows a strong negative relationship between average temperature and station elevation. The plot shows that stations located at higher altitudes tend to record lower average temperatures, indicating a clear influence of altitude on temperature.

Multiple weather indicators can be derived from daily temperature statistics. Average daily temperature values for each station are provided in Appendix Table 14.

Daily minimum, maximum, and mean temperatures are commonly used by MeteoSwiss to classify special day types. For instance, a “summer day” is defined as a day when the

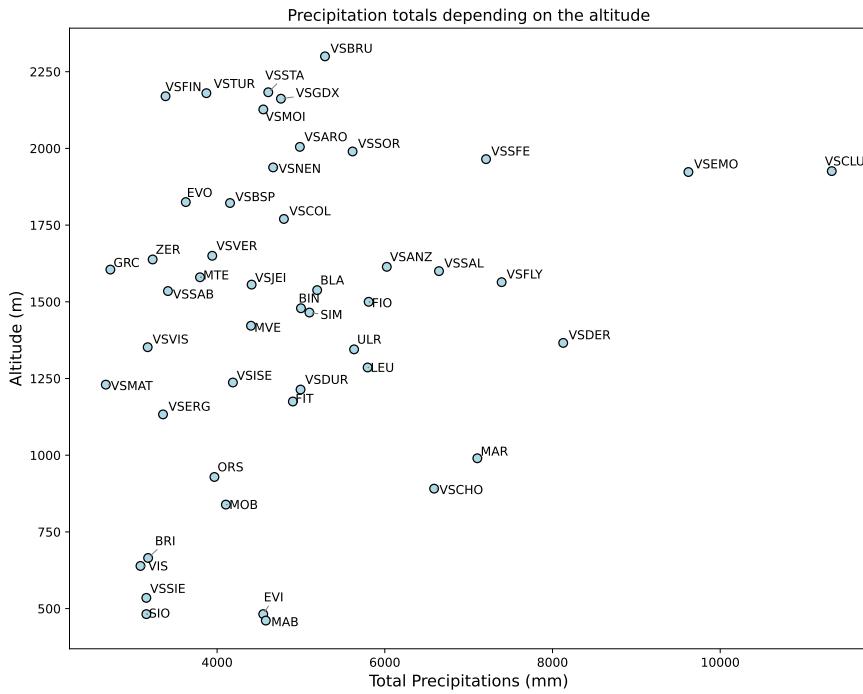


Figure 12: Relationship between total precipitation over the studied period and station altitude (in meters).

maximum temperature exceeds 25°C. Unsurprisingly, Sion leads with a total of 464 summer days over the five-year period, or an average of 92.8 days per year. In contrast, Evolène, a high-altitude station at 1,825 m, recorded only 31 summer days.

Other day types for each station are shown in Figure 14. Interestingly, Sion and Visp show a relatively low number of tropical nights, defined as days when the minimum temperature remains above 20°C. This is somewhat unexpected given the Urban Heat Island (UHI) effect, which typically results in higher night temperatures in urban environments. As expected, alpine stations exhibit significantly more cold-related days. These include heating days (mean temperature $\leq 12^\circ\text{C}$), frost days (minimum temperature $< 0^\circ\text{C}$), and ice days (maximum temperature $< 0^\circ\text{C}$).

A boxplot of relative humidity by station is shown in Figure 15. As illustrated, humidity levels vary significantly depending on the station. The corresponding descriptive statistics are provided in Table 15 in the Appendix.

Evolène appears to be the driest station, while Montagnier records the highest mean humidity levels. Despite these differences, the standard deviation is relatively consistent across stations. Notably, all but four stations reached 100% relative humidity at least once during the observation period.

The average wind direction and speed are displayed in Figure 16. Note that for better readability, the arrows indicate the direction toward which the wind is primarily blowing,

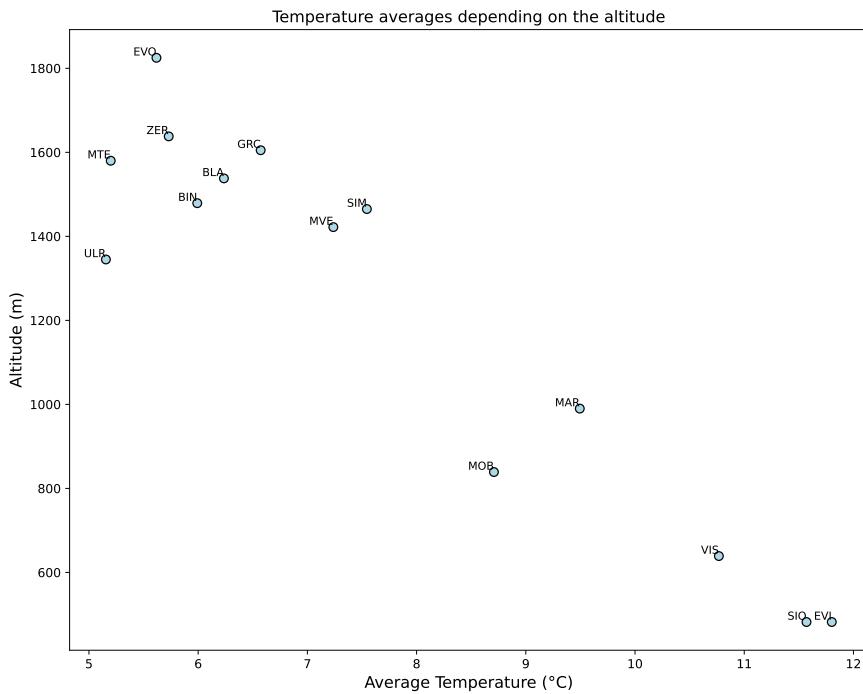


Figure 13: Average temperature over five years plotted against station altitude. A strong negative correlation of -0.94 is observed, indicating a significant altitude effect on temperature.

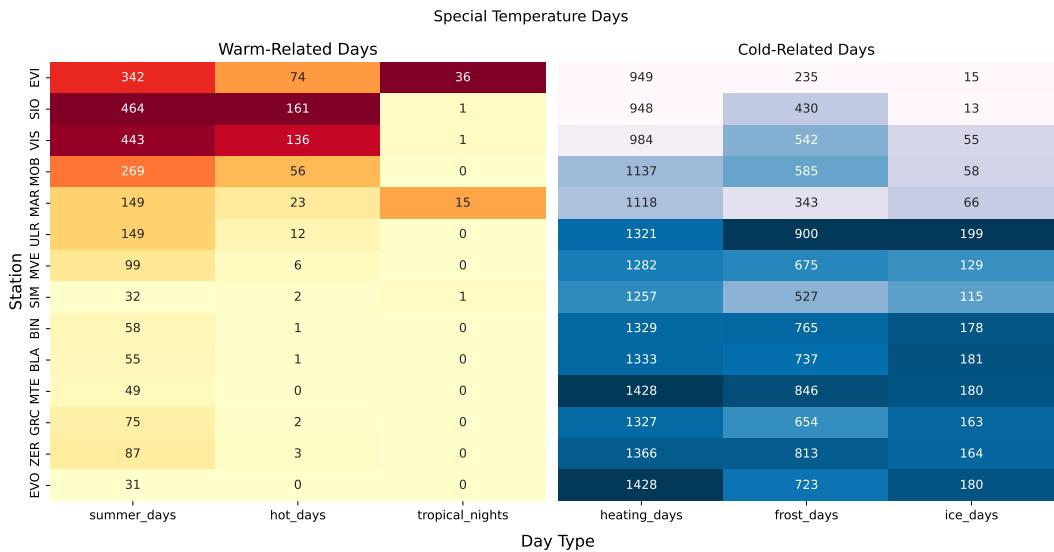


Figure 14: Count of special temperature days for each station over five years. Categories include: summer days ($T_{\max} \geq 25^{\circ}\text{C}$), hot days ($T_{\max} \geq 30^{\circ}\text{C}$), tropical nights ($T_{\min} \geq 20^{\circ}\text{C}$), heating days ($T_{\text{mean}} \leq 12^{\circ}\text{C}$), frost days ($T_{\min} < 0^{\circ}\text{C}$), and ice days ($T_{\max} < 0^{\circ}\text{C}$).

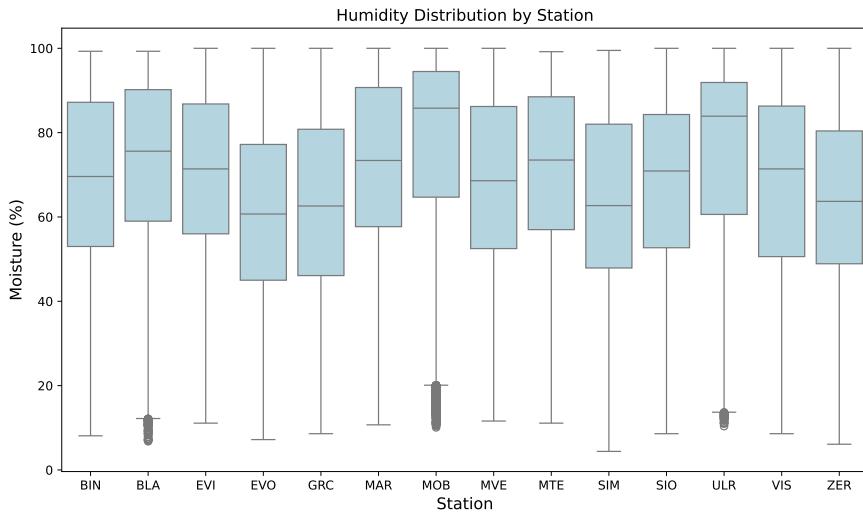


Figure 15: Relative humidity by station, highlighting variability across different locations.

highlighting the predominant wind patterns across the Valais region. In Switzerland, three major wind types dominate: westerly winds, the Foehn (typically blowing from the south), and the Bise (usually coming from the north) [31]. Most stations in Valais show prevailing southerly winds, consistent with the influence of the Foehn. However, some stations display different patterns. For instance, Visp (VIS) and Sion (SIO) exhibit easterly winds. Both are located at lower elevations within the Rhône Valley, which may explain this deviation due to the valley's orientation and topography. In contrast, the station of Mottec (MOT), records a dominant northerly wind. Local residents report that the valley is rarely affected by Foehn events, likely due to the topography of the valley.¹⁴

3.3. Missing values. Stations occasionally have missing values due to various causes, ranging from scheduled maintenance to unexpected equipment failure. This section presents an overview of the missing values, examines their distribution and behavior, and outlines the approach used to address them.

As illustrated in Figure 17, the distribution of missing values across stations is highly uneven. Most stations experience only a few missing periods, while a small subset are full of missing periods. Notably, one outlier is the station of VSBRI in Bricola, which has a total of 1,643 missing periods. A complete list of missing values per station is provided in Appendix 7.

Regarding the variables involved in the missing periods, precipitation emerges as the most affected. Even when excluding the two primary contributors to missing data, *VSBRI* and *VSTSN*, precipitation still exhibits more missing values than any other variable. This is likely due to the fact that, while nearly all stations record precipitation, not all of them

¹⁴MeteoSwiss Blog, *Foehn dans le Haut-Valais* (French only), available at <https://www.meteosuisse.admin.ch/portrait/meteosuisse-blog/fr/2023/10/foehnhautvalais.html>, accessed May 22, 2025.

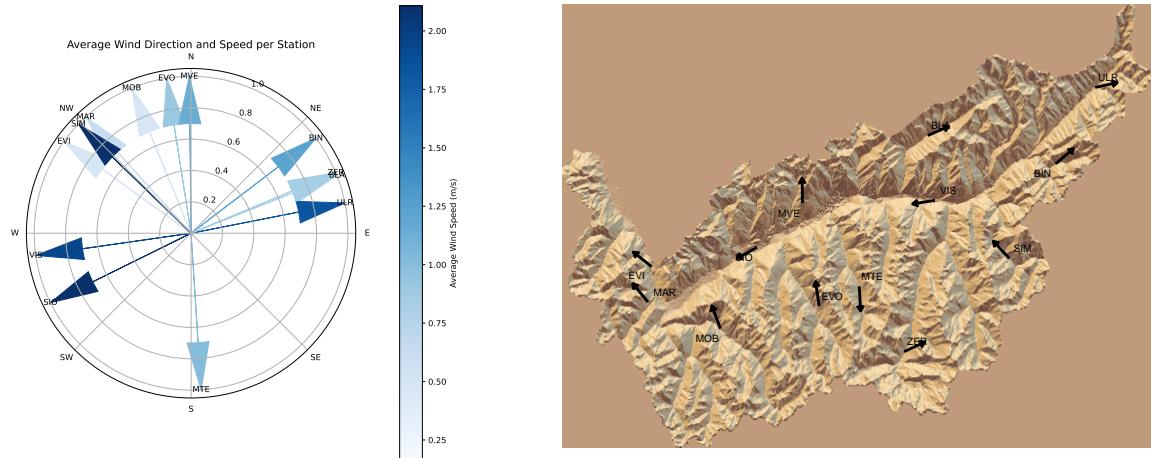


Figure 16: (Right): Average wind direction and speed per station. Arrows indicate the direction the wind is blowing toward, and color represents the average wind speed. (Left): Wind direction visualized over terrain. While displaying direction where the wind blows is at odds to meteorological standards, this convention was adopted for greater clarity.

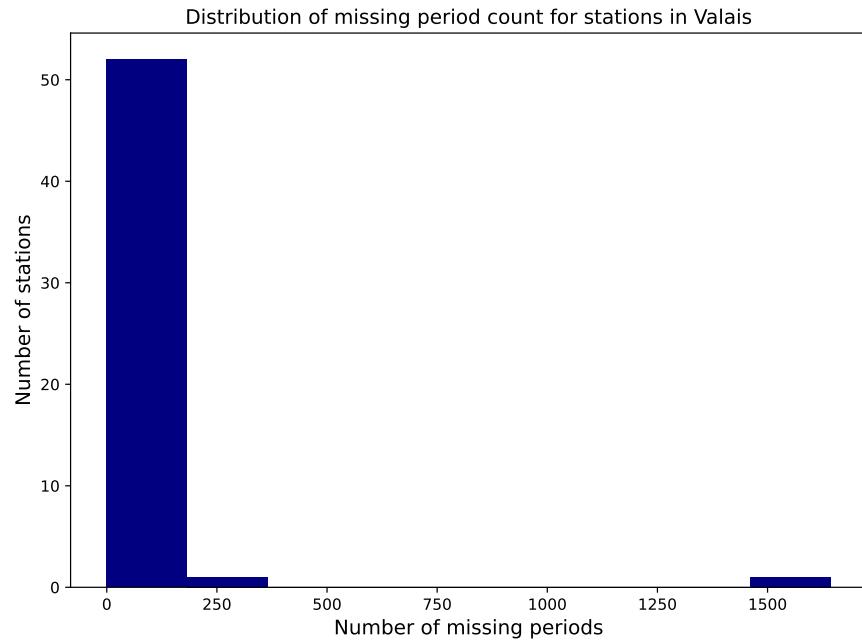


Figure 17: Number of missing periods per station. Most stations show no missing values, while two stations appear as outliers—one with approximately 250 missing periods and another with over 1,600.

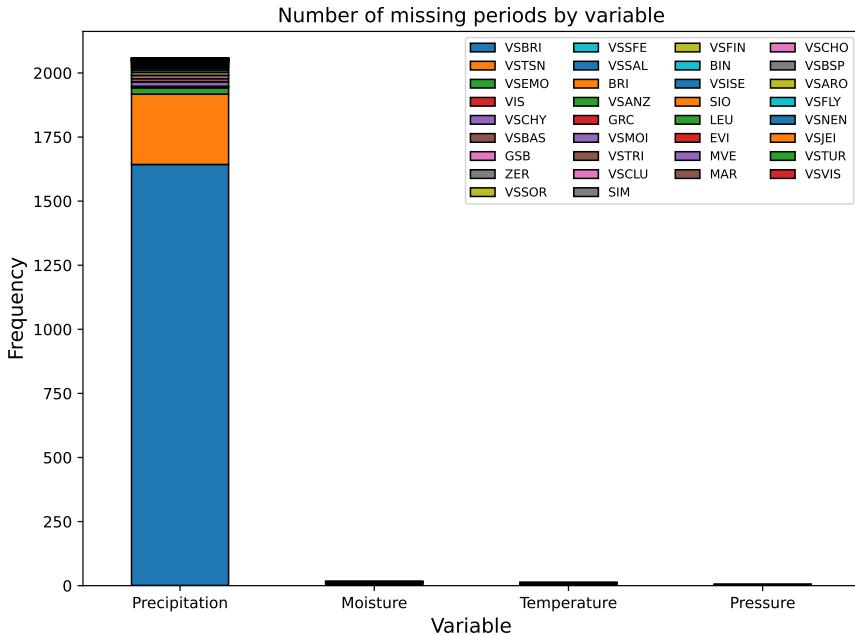


Figure 18: Number of missing periods by variable. The stations in the legend are arranged in decreasing order with respect to precipitation. Wind vectors components are not represented here as they have no missing values.

measure variables such as humidity, temperature, or pressure. Interestingly, the wind vector components (East and North directions) have no missing values.

The duration of missing periods is shown in Figure 19. The vast majority of these periods last less than 12 hours. Given the short duration, such gaps can be reasonably addressed using interpolation techniques.

Additional details concerning the count and distribution of missing values per station are provided in Appendix B.

Figure 20 presents the duration of each missing period across all stations included in the study. The longest interruption occurred at station VSBAS in Baltschiedertal, lasting an exceptional 281 days, 3 hours, and 10 minutes. Remarkably, the same station also holds the second-longest gap, with a duration of 137 days, 5 hours, and 50 minutes. Fortunately, such extended missing periods are rare. The median duration of missing periods is only 40 minutes, indicating that most interruptions are short and can be interpolated.

In conclusion, although some missing values are present in the dataset, the majority are short in duration and occur infrequently. Only a few stations exhibit missing data to an extent that may affect the reliability of the analysis. The following stations are excluded: Bricola (VSBRI), Baltschiedertal (VSBAS), Tsanfleuron (VSTSN), Trient (VSTRI), Champéry (VSCHY), Col du Grand St-Bernard (GSB). While this may seem like a loss of valuable data, it is mitigated by the dense network of meteorological stations in Valais, where

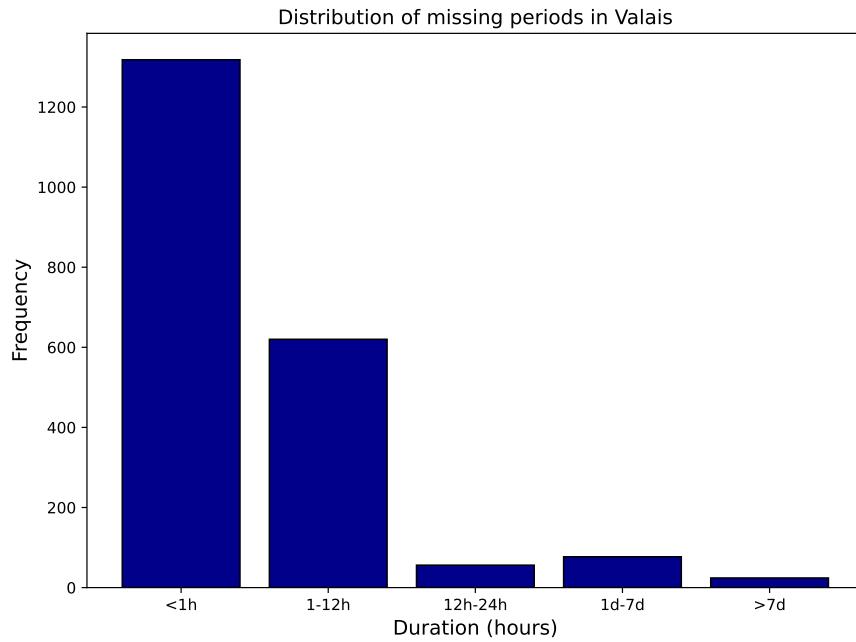


Figure 19: Distribution of missing period durations. With a temporal resolution of ten minutes, one hour corresponds to six consecutive missing values.

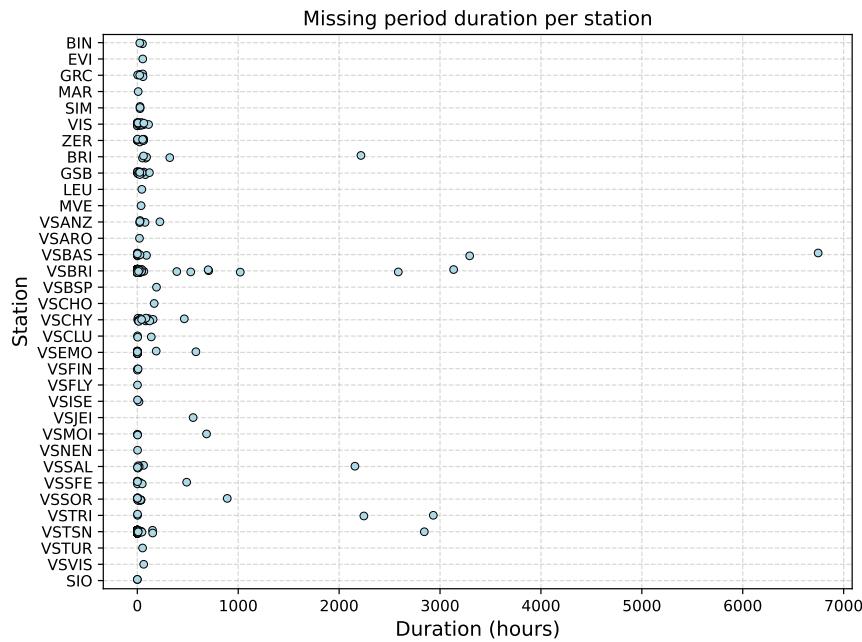


Figure 20: Duration (in hours) of missing periods across all stations.

the average area per station is 92 km^2 , significantly lower than the Swiss national average of 162 km^2 [69].

To address the remaining missing values, interpolation will be performed using Kriging. This geostatistical method is particularly suitable due to its ability to account for spatial autocorrelation, providing robust estimations even in complex topographies like those found in the Alpine region.

4. Data Preparation and Modeling. This section provides an overview of the steps involved in preparing the dataset for predictive modeling. The primary objective is to forecast rainfall levels, which serve as the dependent variable in this study. Specifically, the aim is to predict precipitation levels or the likelihood of precipitation in the future. The process begins with the treatment of missing values, discussed in Section 4.1, where Kriging is employed to interpolate gaps in the dataset, as well as to infer parameters for stations that do not record specific variables. Following interpolation, a series of data transformations are applied, as described in Section 4.2. These procedures standardize, normalize, and structure the dataset, ensuring that temporal dependencies and input consistency are preserved across all samples.

Finally, the predictive model is introduced in Section 4.3, using the `Keras` library in `Python`. The model is trained on the processed dataset to generate rainfall predictions at multiple forecast horizons, specifically using Kolmogorov-Arnold Networks.

4.1. Missing value interpolation. This section outlines the workflow used to interpolate both missing values and unmeasured variables, as well as the procedures used to evaluate the quality of these interpolations.

In the dataset, certain stations do not record all meteorological variables. In fact, out of 54 stations, only 12 record every single variable. To ensure a complete dataset suitable for modeling, these variables are estimated for all timestamps at those locations using Kriging-based interpolation. To reduce the risk of introducing bias into the predictive model, an additional binary column is created for each variable, indicating whether a given value was observed (0) or interpolated (1). This allows the model to differentiate between actual measurements and estimated values during training.

The choice of interpolation method is crucial, as poor estimation introduce errors into the forecasting model. The use of Kriging and the choice of the External Drift variable is supported by relevant studies. Based on the characteristics of each variable, an appropriate parametric form for spatial dependence is selected. In addition, estimations are constrained to avoid implausible interpolated values. For instance, moisture is bounded between 0 and 100 percent, and any predictions falling outside this range are clipped accordingly.

Once model performance is satisfactory, the full interpolation is performed. For each variable, any timestamp with a missing or unmeasured value is estimated using the fitted Kriging model. This is repeated for every relevant station and timestamp. Although the process is parallelized to reduce computational time, it remains resource-intensive. This is due in part to the use of automatic selection of spatial bins instead of using fixed ones.¹⁵

4.1.1. Kriging Model Fitting. The interpolation of meteorological variables was carried out using Kriging with External Drift (KED). Precipitation, temperature, pressure, and hu-

¹⁵For an explanation of spatial bins, see Section 1.4

midity were all interpolated using elevation as the external drift. A comparative study by Tobin et al. [69] focusing on temperature and precipitation found that KED with elevation consistently produced the lowest interpolation bias. Similarly, Rosillon et al. [63] demonstrated that elevation is the most effective drift variable for humidity interpolation.¹⁶.

Wind data, by contrast, is significantly more challenging to interpolate due to its turbulent nature. In this case, a wind atlas is used as the external drift¹⁷. It is important to note that the resulting interpolated wind values should be interpreted as general estimates of regional wind tendencies.

4.1.2. Leave-one-out cross-validation. To evaluate the accuracy of the Kriging interpolation, Leave-One-Out Cross-Validation (LOOCV) is performed. This method consists of systematically removing the observed value for a given variable at a specific station and timestamp, then interpolating that value using the remaining stations as inputs. The process is repeated for each station, and to account for seasonal and annual variability, this procedure is performed across 100 randomly selected timestamps distributed throughout the dataset.

This allows us to determine the quality of the interpolation procedure. For each iteration, the predicted value is compared to the true observed value, and the difference between them is used to compute error metrics. Two standard performance metrics are used, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). These metrics help the identify the most suitable interpolation method for each variable before applying it across the full dataset.¹⁸

In conclusion, the interpolation procedure yields promising results. During the modeling phase, different strategies will be explored to assess the impact interpolation has on forecast accuracy. A key consideration will be to determine whether it is more effective to use only a smaller subset of stations that record all variables directly, or to include all available stations, even if many only record precipitation and rely on interpolation for the remaining variables. Since precipitation is the target variable in this study, retaining stations with direct rainfall observations, even if other variables are estimated, may still offer better precision.

Another important question concerns the sequencing of interpolation and forecasting. Should a model be trained on a smaller subset of fully observed stations and later generalized through interpolation? Or is it preferable to first interpolate all missing variables across a wider network of stations and then train the model using a larger but mostly estimated dataset? The second approach might reduce spatial bias in the final forecast map, but could introduce noise depending on the quality of interpolated inputs.

4.2. Data Transformation. In this section, the various data transformations applied to prepare the dataset for forecasting tasks are presented. The objective is to evaluate whether

¹⁶ Alternative methods exist. For temperature, MeteoSwiss employs an approach proposed by Frei [21], but KED with elevation is preferred here for consistency across variables. For precipitation, the CombiPrecip product developed by MeteoSwiss uses radar as the external drift [66], but such data are unavailable in this study. Finally, while pressure can also be interpolated using temperature as a covariate due to their strong correlation, both approaches yield comparable results, and elevation is retained as the drift.

¹⁷ A wind atlas provides average wind speeds across a spatial grid and serves as a reference in climate studies.

¹⁸ MAE and RMSE are defined as follows: $\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$, $\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$, where \hat{y}_i is the interpolated value, y_i is the corresponding observed value, and n is the total number of interpolated samples, computed as 100 times the number of stations with valid data for the variable in question.

Kolmogorov-Arnold Networks (KANs) are suitable for the problem of rainfall forecasting.

4.2.1. Undersampling. A challenge in rainfall forecasting arises from the highly imbalanced nature of the data. Although the dataset comprises over 263,000 observations¹⁹, approximately 90% represent time steps with no rainfall. This significant class imbalance between dry and rainy events can lead to poor model performance, particularly in classification tasks. Specifically, models may default to always predicting the majority class (no rain), resulting in relatively low error metrics but with poor forecasting ability.

To address this issue, an undersampling strategy is employed. The approach involves retaining only a subset of the majority class (dry observations) to achieve a more balanced dataset. In the training set, only 20% of observations containing no rainfall are retained. This ensures that a diverse set of weather conditions is represented while ensuring a better balance.

It is acknowledged that undersampling leads to the removal of data, which may impact models with less observations. However, due to the large size of the dataset, this loss is acceptable.

4.2.2. Pivoting. The model is designed to generalize across all weather stations. At each time step t , the model outputs 48 forecasts, one for each station. To enable this, each time step must include the observations of all weather variables for all stations. The original dataset is stored in long format, where each row corresponds to a timestamp for a specific station. This format is chosen for reasons of storage efficiency and processing speed, as increasing the number of columns generally leads to larger file sizes, whereas additional rows do not have the same impact.

To prepare the data for model input, a transformation to a wide format is required. This is achieved by pivoting the dataset so that each row corresponds to a single timestamp, and the columns contain the variables for all stations. The input and target can be expressed as:

$$(4.1) \quad \begin{aligned} y_t &= [P_{t,1}, \dots, P_{t,48}], \\ X_t &= [P_{t,i}, HM_{t,i}, TM_{t,i}, AP_{t,i}, U_{t,i}, V_{t,i}] \quad \text{for } i \in [1, 48], \end{aligned}$$

where y_t is the vector of precipitation values at time t for all 48 stations, and X_t contains the corresponding weather variables for each station.

An alternative approach would be to use one input per station and include the station ID as an encoded feature. However, pivoting the data in this way allows the model to exploit spatial correlations between stations, despite the increased memory requirements.

4.3. Model choice. This section details the construction of predictive models based on the weather dataset. The objective is to develop two distinct types of predictive outputs, that address different forecasting needs.

The first weather product is more akin to a classic regression task, returning the occurrence of rain within the next N hours. This formulation is similar to logistic regression and aims to estimate the probability of rain occurrence within a specified future window, using current weather values. Here, N denotes the duration of the window over which rainfall is aggregated.

¹⁹With five years of 10-minute interval observations, the total number of data points is 263,016.

The second product is a standard precipitation forecast, outputting how much precipitation is expected at a future time. This task involves time series regression, where past meteorological observations are used to predict the amount of rainfall at various forecast lead times.

4.3.1. Model Architecture. Unlike traditional time series forecasting methods, neural networks operate by ingesting windows of past observations and learning to identify patterns in the data.

The model architecture can be understood as a progressive task, with all forecasting models in this study sharing a common structure. First, future precipitation values can be aggregated over a defined time window to estimate the likelihood of rain occurrence. If this can be accurately modeled, the next step is to predict the actual amount of rainfall at a single future time step. Once single-step predictions are established, the model can be used recursively to produce forecasts over extended horizons.

This approach from aggregated event prediction, to one step ahead forecasting, and finally to recursive multi-step forecasting is the basis of the modeling framework described here.

Rolling-Windows. To train each model, the time series is segmented into overlapping input-output windows. Each window comprises a sequence of H past observations used as input, and a corresponding target sequence of F future steps. The model is trained to forecast future values based on patterns learned from recent history.

Figure 21 shows an example of such a window. This example is for precipitation for one station. The same window is retrieved for all stations and for all weather variables.

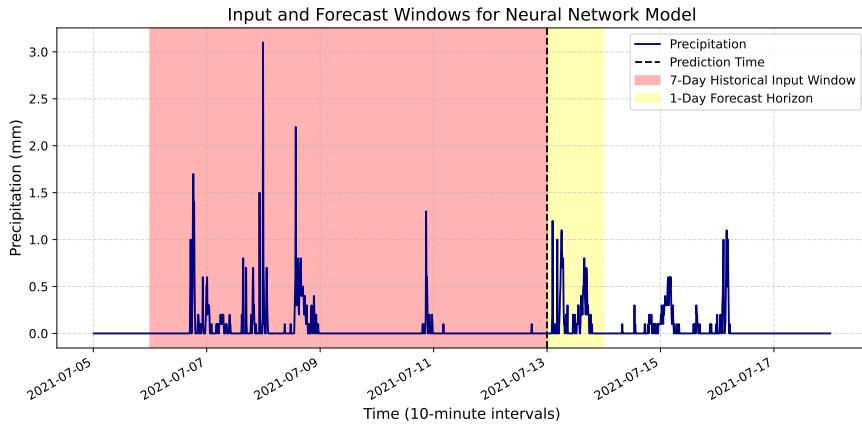


Figure 21: Example of an input-output window for the neural network. The red region represents the historical input (X) and the yellow region the forecast horizon (y). Note that the forecast horizon is only selected for our dependent variable P .

Aggregated Precipitation Classification. To predict the likelihood of precipitation within an aggregated forecast horizon, the total precipitation over the next N time steps is computed for each station. For a given station i , this aggregated rainfall is defined as:

$$P_{t,i} = \sum_{j=0}^N P_{t+j,i}$$

where $P_{t,i}$ denotes the total precipitation received by station i between time steps t and $t+N$.

The occurrence of precipitation is then formulated as a binary classification task, where the target label is defined as:

$$\text{Rain}_{t,i} = \begin{cases} 1 & \text{if } P_{t,i} > 0 \\ 0 & \text{if } P_{t,i} = 0 \end{cases}$$

here, the model is trained to estimate the probability of rain, denoted as $PP_{t,i}$, which represents the probability of rain occurring at station i within the next N time steps.

While the model is structured as a binary classifier, it can be easily adapted to perform regression instead. For instance, it could be modified to predict the total or average amount of precipitation over the window. However, this approach does not provide information about the distribution or variability of precipitation within the forecast window.

Single-Step Forecasting. To forecast multiple time steps into the future, the approach begins by training the model to perform single-step forecasting. The same model is then applied recursively to generate predictions over longer horizons.

The architecture used here is similar to the binary classifier, but the forecast horizon is set to one. The target variable becomes the precipitation at the next time step for station i , written as $P_{t+1,i}$. The model is trained by minimizing a loss function that quantifies the error between predicted and observed precipitation. The task is formally defined as:

$$\hat{P}_{t+1,i} = f_\theta(P_{a,i} + HM_{a,i} + TM_{a,i} + AP_{a,i} + U_{a,i} + V_{a,i} + z_i + x_i + y_i), \quad \text{for } a \in [t-H+1, t],$$

where $\hat{P}_{t+1,i}$ is the predicted precipitation for station i at time $t+1$, and f_θ is the neural network model parameterized by weights θ . The inputs span a history window of length H , capturing the most recent H time steps of each weather variable: precipitation (P), humidity (HM), temperature (TM), pressure (AP), and wind components (U, V).

The static features—altitude (z_i) and spatial coordinates (x_i, y_i)—are specific to each station and do not vary over time.

The model outputs a prediction for one station at a time. Every forecast benefits from inputs from all other stations.²⁰

Forecasting for Zero-inflated problems. One issue arising from the high sparsity is that the distribution of precipitation amounts is heavily skewed towards zero. The model therefore tends to underestimate precipitation. And often returns low values no matter if the model saw rain or not. To solve this issue, the following workflow is implemented. First, for a specific time t , the classifier returns whether for each station there will be rain at all for the

²⁰Access to data from multiple stations enables the model to capture cross-station correlations. For instance, if one station consistently receives precipitation a few hours before another, this temporal pattern can be learned by the model.

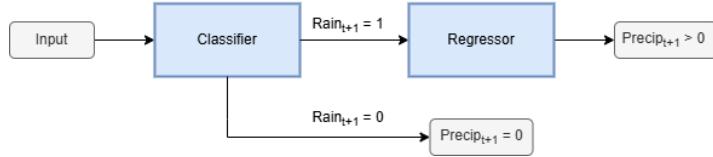


Figure 22: Enter Caption

next time step. If no rain is expected, the forecast will be exactly zero. If rain is expected, the regression model is used to produce forecasts. the regression model is specifically trained only on observations containing rain, ensuring a better forecast of rain events. Figure 22 displays the flowchart of forecasting used.

Multi-step Recursive Forecasts. To extend predictions beyond a single time step, a common approach is to apply recursive forecasting. In this setting, a model trained for single-step forecasting is used iteratively: the prediction at time $t + 1$ is fed back as input to predict $t + 2$, and this process continues until the desired forecast horizon $t + N$ is reached.

Figure 23 illustrates the recursive forecasting process. The historical input window is shown in red, and the predicted outputs in yellow. As each prediction is made, it is appended to the input sequence to form the basis for the next forecast step.

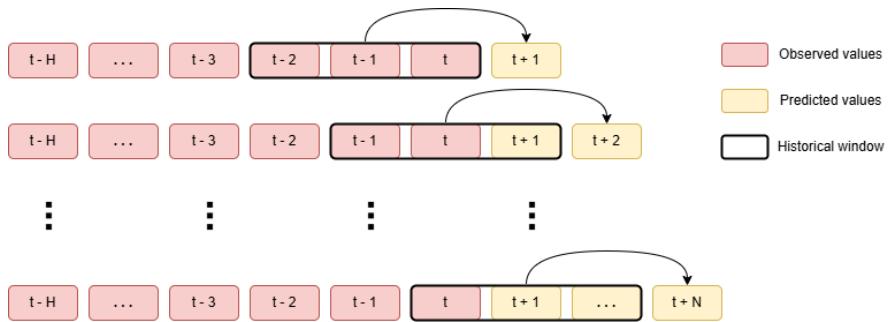


Figure 23: Illustration of recursive forecasting: predictions are made one step at a time using previously predicted values as inputs for the next step.

Recursive forecasting uses the same model repeatedly over the prediction window, allowing it to produce forecasts at further time-steps. This method has a low computational cost, and is flexible, avoiding the need to retrain the model for different forecast horizons. However, recursive forecasting has one notable limitation, it is susceptible to error accumulation. Mistakes made during early prediction steps are part of the input, potentially snowballing into larger inaccuracies over time.

An alternative strategy is direct multi-output forecasting, where a model is trained to predict all future steps simultaneously. Such models are more complex, one of the motivation of this study to find a lightweight model motivates the decision to use recursive models instead of multi-output forecasting.

4.3.2. Model Summary. The forecasting models proposed in this study combine Long Short-Term Memory (LSTM) layers with KAN layers in a hybrid architecture. Pure KAN-based models were initially explored. However, they were unable to produce meaningful forecasts by themselves. As such, a hybrid LSTM-KAN model was adopted to leverage the temporal modeling strength of LSTMs together with the simplicity and interpretability of KANs. The results of the purely KAN-based models are provided in the Appendix, with Figure 31.

Binary Classification Model. Figure 24 displays the architecture used for the binary classification task, where the model predicts the likelihood of precipitation occurring within the next 6 upcoming observations, while the input comprises the previous 36 observations.²¹ The model has a batch size of 256 and uses 10 epochs. The learning rate is set to 1×10^{-4} , this hyperparameter is the step size, determining how fast the model converges.²²

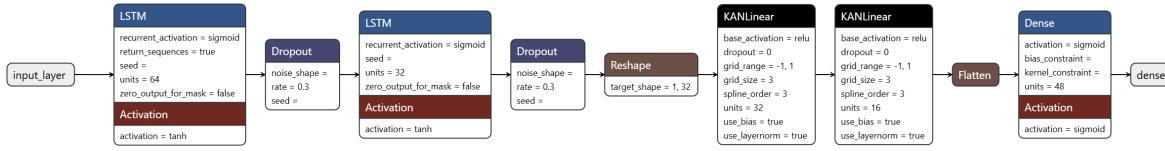


Figure 24: Architecture of the binary classification model. It consists of an LSTM block, a KAN block, and a final Dense layer with sigmoid activation, producing one probability estimate per station.

The model architecture begins with two stacked LSTM layers, each followed by a Dropout layer to reduce overfitting by randomly omitting units during training. The data is then reshaped to match the input requirements of the KAN layers. Two KANLinear blocks follow, processing the features through non-linear transformations with spline-based activations, characteristic of Kolmogorov-Arnold Networks. Finally, a Dense output layer with a sigmoid activation function maps the processed features to probability values in the range $[0, 1]$. The output has a dimensionality of 48, corresponding to the number of weather stations, producing one forecast for each station at the same time.

The model is trained using the Binary Cross Entropy (BCE) loss, this metric calculates the difference between predicted probabilities and the actual labels. Formally:

$$BCE = -\frac{1}{n} \sum_{i=1}^n (y_i \cdot \log(\hat{p}_i) + (1 - y_i) \cdot \log(1 - \hat{p}_i)),$$

²¹With 10-minute intervals between observations, the forecast horizon corresponds to 1 hour, and the historical input spans 6 hours.

²²The input passes through two LSTM layers with 64 and 32 units respectively. Dropout layers with a rate of 0.3 follow each LSTM block to reduce overfitting by randomly deactivating 30% of the nodes during training. The output is reshaped from 2D to a 1D vector to match the input dimensions expected by the KANLinear layer. These KAN layers approximate functions using splines over a grid size of 3, with splines of the 3rd degree. The first KANLinear layer outputs 32 units, followed by a second with 16 units. A Flatten layer is applied to convert the output into a flat vector before passing it into a Dense layer.

where n is the number of observations, the predicted probability is written as \hat{p}_i , and y_i represents the actual label, 0 or 1.

One-Step Forecasting Model. The one-step-ahead forecasting model shares much of the same architecture as the classification model, with one key modification in the output layer. The task here is to directly predict the total precipitation occurring over the next 10-minute interval (i.e., a single future time step). As before, the model takes the previous 36 observations as input, a batch size of 256, 10 epochs and a 1×10^{-4} learning rate. Mean Square Error (MSE) is used as a loss function.²³

Instead of returning a vector of 48 values, the target to predict is the precipitation for a single station. All other layers remain unchanged compared to the previous model. The summary of the model can be seen in Figure 25.

As mentioned previously, the model is only trained on non-zero precipitation observations.

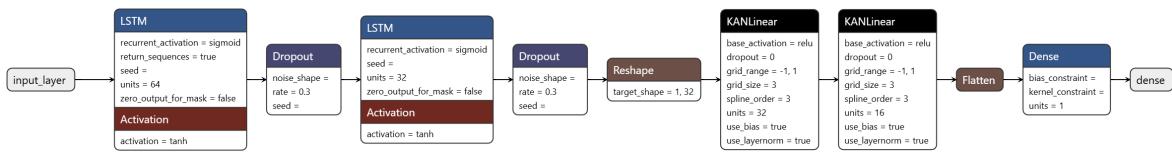


Figure 25: Architecture of the one-step-ahead forecasting model. It includes an LSTM block and a KAN block, with a final Dense layer producing rain amount predictions for one station.

Recursive Forecasting. Recursive forecasting builds upon the classifier model as well as the trained one-step-ahead model without requiring any retraining. For each timestamp, the occurrence of rain is classified. In the case of a positive rain classification, a prediction will be made for the next time step. Otherwise the amount of rain forecasted is set to zero. Each output is appended to the input sequence, while the other variables are set to the last known measurement. The process is repeated until the forecast horizon is reached. To evaluate performance over time, the Mean Absolute Error (MAE) is computed at each forecast step, providing insight into the error accumulation as time passes.

4.3.3. Model Metrics. In binary classification tasks, model performance is commonly assessed using the *confusion matrix*. This matrix compares predicted labels with actual labels and categorizes predictions into four groups: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

A true positive refers to a correctly predicted positive case, while a true negative corresponds to a correctly predicted negative case. A false positive occurs when a negative instance is misclassified as positive, and a false negative when a positive instance is predicted as negative.

Several performance metrics can be derived from the confusion matrix. Accuracy measures the proportion of correctly classified instances overall. Precision reflects the proportion of true positives among all instances predicted as positive. Sensitivity (also known as recall) measures

²³MSE is defined as follows: $\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$, where n is the number of observations \hat{y} the predicted value and y the actual value.

the proportion of true positives among all actual positives, whereas Specificity captures the proportion of true negatives among all actual negatives. These can be combined into the F_1 score, which represents the harmonic mean of precision and sensitivity. The metrics can be expressed formally as:

$$(4.2) \quad \begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\ \text{Precision} &= \frac{TP}{TP + FP} \\ \text{Sensitivity} &= \frac{TP}{TP + FN} \\ \text{Specificity} &= \frac{TN}{TN + FP} \\ F_1 &= 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \end{aligned}$$

Another commonly used evaluation tool is the Receiver Operating Characteristic (ROC) curve, which plots sensitivity against specificity as the classification threshold varies between 0 and 1. A model that makes random predictions yields a diagonal ROC curve. The Area Under the Curve (AUC) quantifies the ROC curve's performance.²⁴

Forecasting tasks use several metrics. In addition to MAE and RMSE which have been detailed in Section 4.1, R-squared (R^2) is used. The R-square metric used here differs slightly in calculation from the one used in statistical regressions, but serves the same purpose. R-square calculates the square of the correlation between predicted and actual values. Formally, it can be expressed as: $R^2 = \text{cor}^2(y, \hat{y})$, where \hat{y} the predicted value and y the actual value.

5. Summary and discussion of the results..

5.1. Kriging Results. Overall, the interpolation results are satisfactory. The dependent variable, precipitation, shows particularly strong performance, providing confidence in the quality of the estimates and their suitability for modeling.

Error metrics for all variables are presented in Table 3. As expected, the quality of interpolation varies by variable. Wind vectors, which represent a potential source of uncertainty due to their turbulent nature, still yield acceptable error values. However, their contribution to the model remains uncertain. To assess their utility, we will evaluate model performance both with and without the inclusion of wind variables.

5.1.1. Interpolation Visualizations. The quality of spatial interpolation is best evaluated through the use of maps. These visualizations provide an good understanding of spatial patterns and estimation accuracy. A similar approach will be adopted for visualizing forecast outputs, as one of the primary goals of weather forecasting is to communicate predictions to a wide audience. All current interpolation results are displayed in Figure 26

²⁴The AUC measures the area under the ROC curve and ranges from 0.5 (random guessing) to 1 (perfect classification). Higher values indicate better performance.

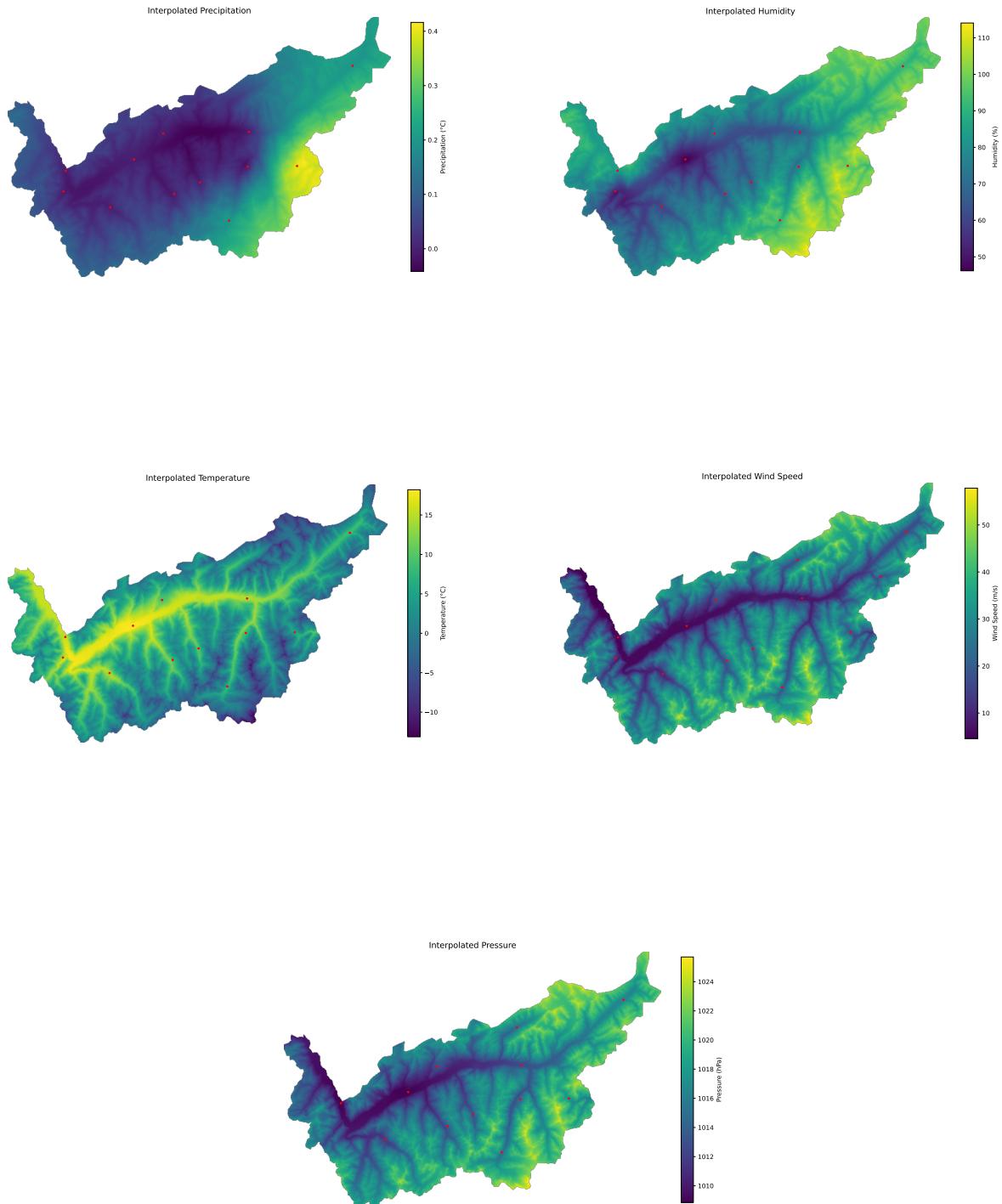


Figure 26: Visualization of interpolated meteorological variables across the Valais region on May 10, 2025. In red are the stations that record observations.

Metric	MAE	RMSE	Drift	Unit
P	0.015	0.059	Elevation	mm
TM	1.594	2.209	Elevation	°C
AP	0.726	1.016	Elevation	hPa
HM	9.93	12.966	Elevation	%
U	1.506	2.116	Avg. wind speed	–
V	1.740	2.443	Avg. wind speed	–

Table 3: Error metrics from Leave-One-Out Cross-Validation (LOOCV) across 100 random timestamps. Reported metrics are Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The external drift variable used in each interpolation is also indicated.

5.2. Model Results. This section explores the results of the models described in Section 4.3, starting with the classifier, and followed by the one-step-ahead model. Lastly, the recursive forecasting capabilities will be evaluated. One should note that the models are not compared to benchmark models. The quality is assessed based on error metrics.

Additional models using different aggregation schemes and time resolutions were also trained. These are not discussed in the main text but are available in the Appendix, with Figure 18 showing the test metrics.

5.2.1. Binary Classifier Model. The model was trained and its architecture, along with the weights, was saved for future prediction use. In the final epoch, the model achieved an AUC of 0.931 and a Binary Cross-Entropy loss of 0.260. On the validation set, the AUC reached 0.951 with a loss of 0.148. These results indicate that the model converged effectively and identified meaningful patterns in the data windows.

Test metrics are summarized in Table 4. The consistently high performance across all metrics, and especially the small AUC gap between training and test sets, suggest a good generalization capacity and the absence of overfitting. Notably, sensitivity, a crucial metric in precipitation forecasting reached 0.692. This means that 69% of rain events were correctly identified. Given the class imbalance in the dataset, this result is encouraging and demonstrates that the model does not overly favor the dominant no precipitation class, supporting the quality of the undersampling process.

Metric	Score
Precision	0.779
Accuracy	0.932
Sensitivity (Recall)	0.692
F1 Score	0.733
AUC	0.951

Table 4: Goodness-of-fit metrics for the binary classifier on the test set. The small difference in AUC compared to training confirms the model’s generalization.

The score of 69% leaves room for improvement. More aggressive undersampling might help, though it risks reducing the number of available data points too much. Alternatively, modifying the loss function to weigh rain events more heavily could improve recall, though this could compromise the accurate prediction of dry events. Overall, the model shows a strong balance between recall for rain and the preservation of dry event accuracy.

The ROC curve (Figure 27a) further illustrates the model's capability, the AUC of 0.951 is very close to a perfect classifier (AUC=1). Additionally, the normalized confusion matrix (Figure 27b) echoes the sensitivity score. It reveals that 97% of dry events and 69% of rainy events were classified correctly. While 30% of rain events were misclassified as dry, this is still within an acceptable range and does not undermine the overall utility of the model.

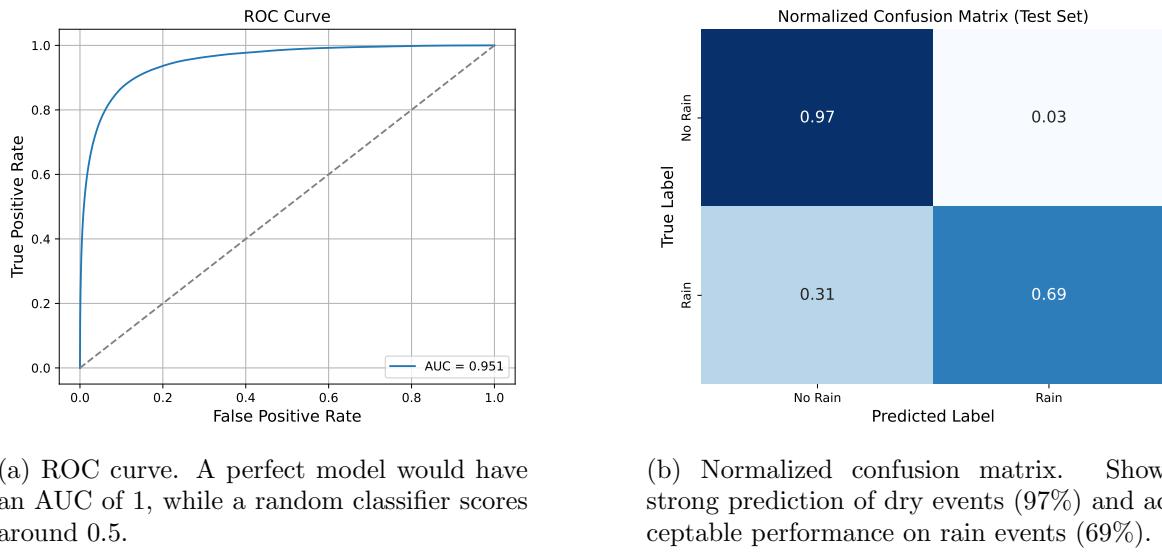
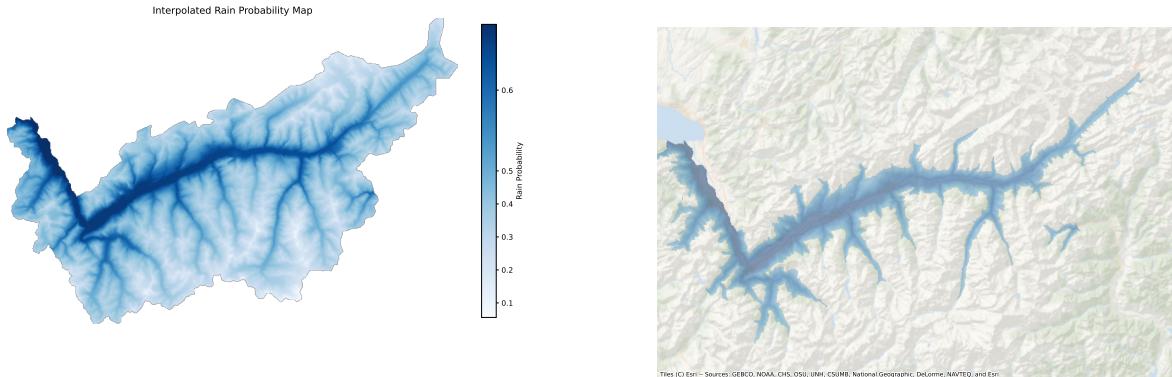


Figure 27: Performance visualizations for the binary classifier model.

Lastly, predictions can be visualized by interpolating on a map with a 1 km resolution. This is a useful tool for understanding and communicating where rain is likely to occur, making it more accessible to the public. However, it is important to stress that this visualization does not convey the quality or reliability of the forecast.

Figure 28a shows the predicted rain probabilities using elevation-based kriging. In Figure 28b, the same probabilities are overlaid on a geographic basemap, and filtered to only include areas where the probability of rain was above 50%.

5.2.2. One-step-ahead Model. Once trained, the model was saved for use in future predictions and recursive forecasting. In the final epoch, the model finished training with a validation loss of 0.027 and a MAE of 0.109, suggesting that the forecasts are close to the actual values. Test set performance metrics are presented in Table 5. While the low error values (MAE, MSE, RMSE) indicate that predictions are numerically close to the true values, the r-square coefficient $R^2 = 0.252$ suggests only a weak linear correlation between predicted



(a) Interpolated rain probability map using elevation-based Kriging. Darker areas indicate higher likelihood of rain.

(b) Interpolation layered over a basemap, only values classified as rain are kept.

Figure 28: Visualization of predicted rain probability over the canton of Valais. These maps offer a way to interpret the spatial distribution of forecasts, but should not be interpreted as a direct measure of model performance.

and observed outcomes. This indicates that the model may not be learning the predictive patterns in the data.

Metric	Score
RMSE	0.166
MSE	0.028
MAE	0.113
R^2	0.228

Table 5: Goodness-of-fit metrics for the one-step-ahead forecasting model on the test set. Despite low error values, the weak R^2 score indicates poor correlation with the actual values.

Figure 29 provides further insights by plotting predicted values against actual values. Most predictions tend to cluster near the origin, revealing the model's inability to forecast extreme precipitation events. Notably, the predicted maximum is just above 1 mm, while the true maximum in the test set is nearly 3 mm, indicating a strong underestimation bias.

In summary, the model shows limited predictive power. Several factors contribute to this, most notably the skewed distribution of rain events toward low values, which introduces a bias.

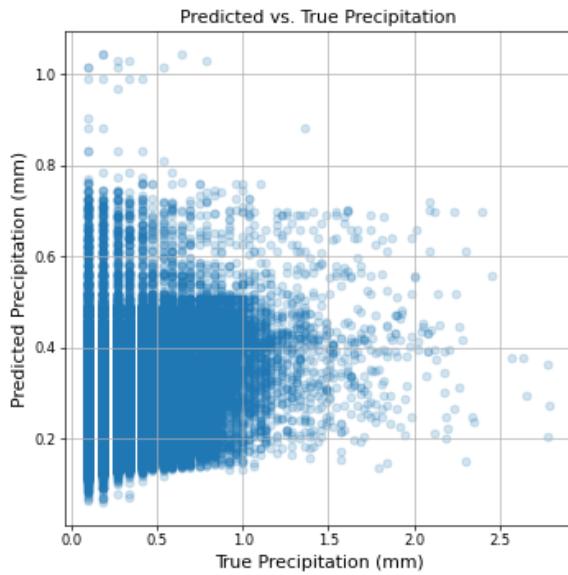


Figure 29: Scatter plot of predicted vs. actual values. The discrepancy in scale reveals that rainfall is consistently underestimated.

However, it is important to highlight that the model is specifically trained and evaluated only on samples containing rain events. Therefore, the reported errors should be interpreted as a reflection of rain prediction accuracy, rather than of general weather forecasting performance.

5.2.3. Multi-step Forecasting. Multi-step forecasts are produced by first applying the binary classifier to predict rain occurrence, and if applicable, passing the results through the regression model to estimate rainfall quantities. To evaluate recursive forecasting performance, a random sample of 100 time windows is selected. Forecasts are then generated recursively from 10 minutes up to 48 hours into the future. The mean absolute error (MAE) is computed for each forecast horizon and plotted in Figure 30. Detailed statistics for the first 24 hours can be found in Table 19 in the Appendix.

The overall error remains surprisingly low across horizons. However, it is important to note that these results include many zero or near-zero rainfall events, which can lead to an overly optimistic assessment of model performance. Despite this, the model maintains stable accuracy over 24 hours, the second day sees higher volatility, with some error spikes.

5.3. Results discussion. The combination of several models and methods allow to create interesting forecasts that may be used to create meaningful forecasts that have real-life utility. Starting with Kriging, the ability to turn single station observations into gridded observations has a lot of potential even by themselves. While most weather variables returned good results, Precipitations and Temperature particularly stand out by their ability to be generalized into a grid, the elevation drift also helps to deal with the complex alpine terrain. Even by themselves, The interpolation of Precipitation into a map shows utility even by themselves. This is

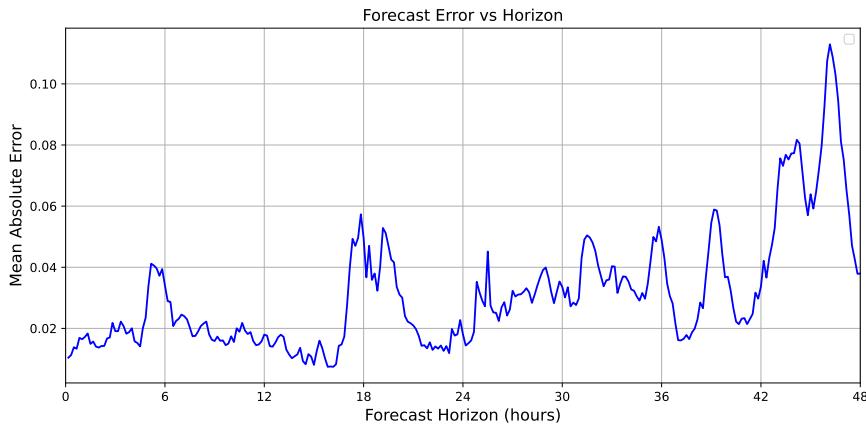


Figure 30: Forecasting error (MAE) as a function of time horizon.

applicable for regions with complex terrain where radars may sometimes struggle, or areas with lower funding. Or even for areas requiring very fine resolution. One could imagine setting up several low cost weather stations in urban areas to better analyze and determine weather effects, such as rain or Urban heat island effect [30]. While not the main focus of this study, Kriging interpolation has taken up a sizable place and shown potential for further focus in future studies.

The use of Kriging has also proved useful in the process of interpolating missing values. Weather stations measurements can be missing for a variety of reasons, from natural disasters to scheduled maintenance, being able to keep an uninterrupted timeline is a welcome addition. Moreover, spatial interpolation can also be used to cross-check values against themselves. Indeed, even if a station measures something, it may not fully reflect the ground truth. An example, our maximum 10-minute rainfall recorded value in Arolla is above the official MeteoSwiss record in Lausanne. Being able to cross-reference data and having an outlier detection pipeline to ensure that outliers due to outside interference are detected and can be interpolated with outside measurements is beneficial.

Interpolated timeseries have enabled two things, first no missing values and secondly, stations that only record precipitation have recordings as well. This has enabled us to create accurate forecasts.

The first model, uses a binary classifier to determine the likelihood of precipitation in the next hour. Metrics show that the model accurately defines when no rain is expected. It is also able to classify rain events in almost 70% of all cases. This does leave us with 30% of unexpected rainy events, indicating that predictions could still be improved. This likelihood generalizes well and the same methodology could be applied to longer timeframes, classifying rain likelihoods in the next days or weeks.

The second trained model, the one-step forecaster returns more difficult results. Indeed, while the error metrics show that forecasts are often close, Further testing show that the model struggles to accurately predict true rainfall values. The poorer performance of the model likely

stems from the unbalanced distribution of observations. Indeed, most rain observations are of a small magnitude, when sudden and powerful rain showers such as those that may be experienced during summer storms happen, the model, primarily trained on lower precipitation amounts returns a lower value than what may actually be observed.

To solve this issue, two models may be created, one solely focusing on everyday precipitations, and another focused on extreme rain events, applying Extreme Value Theorems and other Risk Analytics frameworks. While beyond the scope of this study, the models proposed here lay a groundwork for the first model.

Lastly, the recursive model showed interesting results. Combining both a classifier and a regressor task allows to bypass events containing no rain. While we know that the classifier model has a bias towards False Negatives, it still manages to create forecasts over longer periods of time. Furthermore, the evolution of error as the horizon increases also give us information regarding the

Overall, these models all show potential, while the usage of KAN networks by themselves does not prove effective, the layers still offer an interesting alternative. Benchmark testing has shown that under the right conditions, a KANLinear may prove faster and return less errors than a Dense layer.

6. Limitations and Conclusion.

6.1. Limitations. Importantly especially in terms of Interpolation and overall quality of data, stations bordering the Canton of Valais could have been used. This bordering data could prove useful in subsequent studies to enhance the quality of forecasts for border regions.

Appendix A. Static variables for each station.

Appendix B. Missing periods depending on the interval.

Appendix C. Descriptive Statistics.

Appendix D. Pure KAN model.

Appendix E. Results.

Acknowledgments. The datasets have been provided by MeteoSwiss, the Swiss Federal Office of Meteorology and Climatology.

REFERENCES

- [1] S. ABDULAH, A. H. BAKER, G. BOSILCA, Q. CAO, S. CASTRUCCIO, M. G. GENTON, D. E. KEYES, Z. KHALID, H. LTAIEF, Y. SONG, ET AL., *Boosting earth system model outputs and saving petabytes in their storage using exascale climate emulators*, in SC24: International Conference for High Performance Computing, Networking, Storage and Analysis, IEEE, 2024, pp. 1–12.
- [2] K. ABHISHEK, M. P. SINGH, S. GHOSH, AND A. ANAND, *Weather forecasting model using artificial neural network*, Procedia Technology, 4 (2012), pp. 311–318.
- [3] M. AICHINGER-ROSENBERGER, E. BROCKMANN, L. CROCETTI, B. SOJA, AND G. MOELLER, *Machine learning-based prediction of alpine foehn events using gnss troposphere products: first results for alt-dorf, switzerland*, Atmospheric Measurement Techniques, 15 (2022), pp. 5821–5839.
- [4] M. G. ALTARABICHI, *Dropkan: Regularizing kans by masking post-activations*, arXiv preprint arXiv:2407.13044, (2024).

Table 6: Static variables used for station description

Variable	Explanation	Example
<i>Station name</i>	Full station name	Zermatt
<i>Station</i>	MeteoSwiss station identifier	ZER
<i>WIGOS-ID</i>	WMO station identifier	0-20000-0-06748
<i>Station type</i>	Variables recorded by the station	"Station meteo"
<i>Owner</i>	Owner of the station	"MeteoSuisse, SLF"
<i>Data since</i>	Date of first measurements	01.04.1892
<i>Altitude</i>	Altitude above sea level (m)	1638
<i>Barometer Altitude</i>	Altitude of the barometer (m)	1640
<i>Easting</i>	Easting in CH1903+ / LV95	2624298
<i>Northing</i>	Northing in CH1903+ / LV95	1097574
<i>Latitude</i>	Latitude ($^{\circ}$)	46.029272
<i>Longitude</i>	Longitude ($^{\circ}$)	7.752433
<i>Exposition</i>	Wind exposure direction	"Versant ouest"
<i>Canton</i>	Canton abbreviation	VS
<i>Measures</i>	Variables measured	Temperature, . . . , Wind
<i>Link</i>	Station info link	Link

- [5] S. S. BABOO AND I. K. SHEREEF, *An efficient weather forecasting system using artificial neural network*, International journal of environmental science and development, 1 (2010), p. 321.
- [6] A. Y. BARRERA-ANIMAS, L. O. OYEDELE, M. BILAL, T. D. AKINOSH, J. M. D. DELGADO, AND L. A. AKANBI, *Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting*, Machine Learning with Applications, 7 (2022), p. 100204.
- [7] P. BAUER, A. THORPE, AND G. BRUNET, *The quiet revolution of numerical weather prediction*, Nature, 525 (2015), pp. 47–55.
- [8] K. BI, L. XIE, H. ZHANG, X. CHEN, X. GU, AND Q. TIAN, *Accurate medium-range global weather forecasting with 3d neural networks*, Nature, 619 (2023), pp. 533–538.
- [9] A. D. BODNER, A. S. TEPSICH, J. N. SPOLSKI, AND S. POURTEAU, *Convolutional kolmogorov-arnold networks*, arXiv preprint arXiv:2406.13155, (2024).
- [10] G. E. BOX, G. M. JENKINS, G. C. REINSEL, AND G. M. LJUNG, *Time series analysis: forecasting and control*, John Wiley & Sons, 2015.
- [11] M. BUZZI, *Challenges in operational numerical weather prediction at high resolution in complex terrain*, PhD thesis, ETH Zurich, 2008.
- [12] S.-A. CHEN, C.-L. LI, N. YODER, S. O. ARIK, AND T. PFISTER, *Tsmixer: An all-mlp architecture for time series forecasting*, 2023, <https://arxiv.org/abs/2303.06053>, <https://arxiv.org/abs/2303.06053>.
- [13] S.-Y. CHO, S. LEE, AND H.-G. KIM, *Forecasting vix using interpretable kolmogorov-arnold networks*, arXiv preprint arXiv:2502.00980, (2025).
- [14] CSCS – SWISS NATIONAL SUPERCOMPUTING CENTRE, *"Alps": towards a geographically redundant supercomputing research infrastructure*, Jan. 2024, <https://www.cscs.ch/publications/press-releases/2024/alps-towards-a-geographically-redundant-supercomputing-research-infrastructure>. Press release, accessed 24 Apr 2025.
- [15] C. DE BOOR, *On calculating with b-splines*, Journal of Approximation theory, 6 (1972), pp. 50–62.
- [16] P. S. DE MATTOS NETO, J. F. L. DE OLIVEIRA, D. S. D. O. S. JÚNIOR, H. V. SIQUEIRA, M. H. D. N. MARINHO, AND F. MADEIRO, *A hybrid nonlinear combination system for monthly wind speed forecasting*, IEEE Access, 8 (2020), pp. 191365–191377.
- [17] L. ESPEHOLT, S. AGRAWAL, C. SØNDERBY, M. KUMAR, J. HEEK, C. BROMBERG, C. GAZEN,

Table 7: Missing period count per station

missing Count	station
0	BLA, EVO, FIO, FIT, MAB, MOB, MTE, ORS, ULR, VS-BRU, VSCOL, VSDER, VSDUR, VSERG, VSGDX, VS-MAT, VSSAB, VSSIE, VSSTA, VSVER
1	EVI, LEU, MAR, MVE, VSARO, VSBSP, VSCHO, VSFLY, VSJEI, VSNEN, VSTUR, VSVIS
2	BIN, SIO, VSISE
3	SIM, VSCLU, VSFIN
4	VSMOI, VSTRI
5	GRC, VSANZ
6	BRI
7	VSSAL
8	VSSFE
11	VSSOR
12	GSB, ZER
13	VSBAS
17	VSCHY
22	VIS
25	VSEMO
274	VSTSN
1643	VSBRI

Table 8: Stations with Missing Periods <1h

missing Count	station
1	VSSFE
2	GSB, SIO, VSBAS, ZER
5	VIS
23	VSEMO
239	VSTSN
977	VSBRI

R. CARVER, M. ANDRYCHOWICZ, J. HICKEY, ET AL., *Deep learning for twelve hour precipitation forecasts*, Nature communications, 13 (2022), pp. 1–10.

- [18] FEDERAL OFFICE OF TOPOGRAPHY SWISSTOPO, *Dhm25 / 200m*. <https://www.swisstopo.admin.ch/en/height-model-dhm25-200m>, 2024. Accessed: 30 May 2025.
- [19] J. FISLER, M. KUBE, C. STOCKER, AND E. GRÜTER, *Quality assessment using meteo-cert: The meteoswiss classification procedure for automatic weather stations*, Accessed online: <https://library.wmo.int/idurl/4/55645> (26 May 2025), (2017).

Table 9: Stations with Missing Periods between 1-12h

missing Count	station
1	GRC, MAR, VSFLY, VSISE, VSNEN
2	GSB, VSCLU, VSTRI, ZER
3	VIS, VSFIN, VSMOI, VSSAL
4	VSCHY, VSSFE
6	VSSOR
7	VSBAS
29	VSTSN
609	VSBRI

Table 10: Stations with Missing Periods in 12h-24h

missing Count	station
1	VSANZ, VSARO, VSCHY, VSISE, VSSFE
2	VSSAL
5	VIS
38	VSBRI

- [20] A. FOEHN, J. G. HERNÁNDEZ, B. SCHAEFFLI, AND G. DE CESARE, *Spatial interpolation of precipitation from multiple rain gauge networks and weather radar data for operational applications in alpine catchments*, Journal of Hydrology, 563 (2018), pp. 1092–1110.
- [21] C. FREI, *Interpolation of temperature in a mountainous region using nonlinear profiles and non-euclidean distances.*, International Journal of Climatology, 34 (2014).
- [22] J. GARCÍA HERNÁNDEZ, A. CLAUDE, J. PAREDES ARQUIOLA, B. ROQUIER, AND J.-L. BOIL LAT, *Integrated flood forecasting and management system in a complex catchment area in the alps—implementation of the minerve project in the canton of valais*, River Flow 2014, Special Session on Swiss Competences in River Engineering and Restoration, (2014), pp. 87–97.
- [23] R. GENET AND H. INZIRILLO, *Tkan: Temporal kolmogorov-arnold networks*, arXiv preprint arXiv:2405.07344, (2024).
- [24] U. GERMANN, M. BOSCACCI, L. CLEMENTI, M. GABELLA, A. HERING, M. SARTORI, I. V. SIDERIS, AND B. CALPINI, *Weather radar in complex orography*, Remote Sensing, 14 (2022), p. 503.
- [25] P. GOOVAERTS, *Geostatistics for natural resources evaluation*, Oxford university press, 1997.
- [26] X. HAN, X. ZHANG, Y. WU, Z. ZHANG, AND Z. WU, *Kan4tsf: Are kan and kan-based models effective for time series forecasting?*, arXiv preprint arXiv:2408.11306, (2024).
- [27] K. HE, X. ZHANG, S. REN, AND J. SUN, *Deep residual learning for image recognition*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
- [28] H. HERSBACH, B. BELL, P. BERRISFORD, S. HIRAHARA, A. HORÁNYI, J. MUÑOZ-SABATER, J. NICOLAS, C. PEUBEY, R. RADU, D. SCHEPERS, ET AL., *The era5 global reanalysis*, Quarterly journal of the royal meteorological society, 146 (2020), pp. 1999–2049.
- [29] H. HEWAMALAGE, C. BERGMEIR, AND K. BANDARA, *Recurrent neural networks for time series forecasting: Current status and future directions*, International Journal of Forecasting, 37 (2021), pp. 388–427.
- [30] S. HSU, A. MAVROGIANNI, AND I. HAMILTON, *Comparing spatial interpolation techniques of local urban temperature for heat-related health risk estimation in a subtropical city*, Procedia engineering, 198

Table 11: Stations with Missing Periods between 1d-7d

missing count	station
1	EVI, LEU, MVE, VSCHO, VSCLU, VSSAL, VSSFE, VSTUR, VSVIS
2	BIN, VSBAS
3	SIM, VSANZ
4	BRI, GRC, VSSOR
5	VSTSN
8	GSB, ZER
9	VIS
11	VSCHY
12	VSBRI

Table 12: Stations with Missing Periods >7d

missing count	station
1	VSANZ, VSBSP, VSCHY, VSJEI, VSMOI, VSSAL, VSSFE, VSSOR, VSTSN
2	BRI, VSBAS, VSEMO, VSTRI
7	VSBRI

(2017), pp. 354–365.

- [31] P. JUNGO, S. GOYETTE, AND M. BENISTON, *Daily wind gust speed probabilities over switzerland according to three types of synoptic circulation*, International journal of climatology, 22 (2002), pp. 485–499.
- [32] Z. KAREVAN AND J. A. SUYKENS, *Transductive lstm for time-series prediction: An application to weather forecasting*, Neural Networks, 125 (2020), pp. 1–9.
- [33] R. KEISLER, *Forecasting global weather with graph neural networks*, arXiv preprint arXiv:2202.07575, (2022).
- [34] A. N. KOLMOGOROV, *On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables*, American Mathematical Society, 1961.
- [35] M. KÖPPEN, *On the training of a kolmogorov network*, in Artificial Neural Networks—ICANN 2002: International Conference Madrid, Spain, August 28–30, 2002 Proceedings 12, Springer, 2002, pp. 474–479.
- [36] F. KRISTIANTI, J. DUJARDIN, F. GERBER, H. HUWALD, S. W. HOCH, AND M. LEHNING, *Combining weather station data and short-term lidar deployment to estimate wind energy potential with machine learning: a case study from the swiss alps*, Boundary-Layer Meteorology, 188 (2023), pp. 185–208.
- [37] S. KÜÇÜKDERMENCI, *Design of a cost-effective weather station with rain forecast*, in 2nd International Conference on Scientific and Innovative Studies-ICSIS 2024, 2024.
- [38] M.-J. LAI AND Z. SHEN, *The kolmogorov superposition theorem can break the curse of dimensionality when approximating high dimensional functions*, arXiv preprint arXiv:2112.09963, (2021).
- [39] A. LECLERC, E. KOCH, M. FELDMANN, D. NERINI, AND T. BEUCLER, *Improving predictions of convective storm wind gusts through statistical post-processing of neural weather models*, 2025, <https://arxiv.org/abs/2504.00128>, <https://arxiv.org/abs/2504.00128>.

Table 13: Top 5 Longest Dry Periods

station	name	start date	count
ULR	Ulrichen	2020-03-13 01:50	5424
VSSAB	Saas Balen	2022-02-22 08:40	5314
MAB	Martigny	2020-03-20 18:20	5307
ZER	Zermatt	2022-02-21 19:40	5249
SIO	Sion	2023-01-18 01:30	5183

Table 14: Minimum, Maximum and Mean values over the period

Station	Minimum	Maximum	Mean
BIN	1.61	10.83	5.99
BLA	2.03	10.75	6.24
EVI	7.95	16.25	11.80
EVO	1.91	10.12	5.62
GRC	2.80	11.41	6.57
MAR	6.26	13.78	9.49
MOB	3.70	14.76	8.71
MTE	0.78	10.57	5.20
MVE	2.67	12.05	7.24
SIM	4.31	11.08	7.54
SIO	5.95	17.71	11.57
ULR	-0.90	11.76	5.16
VIS	4.36	16.91	10.77
ZER	1.32	11.40	5.73

- [40] L. LI, R. CARVER, I. LOPEZ-GOMEZ, F. SHA, AND J. ANDERSON, *Generative emulation of weather forecast ensembles with diffusion models*, Science Advances, 10 (2024), p. eadk4489.
- [41] B. LIM AND S. ZOHREN, *Time-series forecasting with deep learning: a survey*, Philosophical Transactions of the Royal Society A, 379 (2021), p. 20200209.
- [42] T. LIU, *awesome-kan: A curated list of resources and papers on kolmogorov-arnold networks (kans)*. <https://github.com/mintisan/awesome-kan>, 2024. Accessed: 2025-04-23.
- [43] Y. LIU, E. RACAH, J. CORREA, A. KHOSROWSHAH, D. LAVERS, K. KUNKEL, M. WEHNER, W. COLLINS, ET AL., *Application of deep convolutional neural networks for detecting extreme weather in climate datasets*, arXiv preprint arXiv:1605.01156, (2016).
- [44] Z. LIU, P. MA, Y. WANG, W. MATUSIK, AND M. TEGMARK, *Kan 2.0: Kolmogorov-arnold networks meet science*, arXiv preprint arXiv:2408.10205, (2024).
- [45] Z. LIU, Y. WANG, S. VAIDYA, F. RUEHLE, J. HALVERSON, M. SOLJAČIĆ, T. Y. HOU, AND M. TEGMARK, *Kan: Kolmogorov-arnold networks*, arXiv preprint arXiv:2404.19756, (2024).
- [46] A. MACDONALD, J. LEE, AND S. SUN, *Qnh: Design and test of a quasi-nonhydrostatic model for mesoscale weather prediction*, Monthly Weather Review, 128 (2000), pp. 1016–1036.

Table 15: Descriptive statistics for humidity by station

station	count	mean	std	min	25%	50%	75%	max
EVO	262944.00	61.37	20.43	7.20	45.00	60.70	77.20	100.00
GRC	262800.00	63.39	21.25	8.60	46.10	62.60	80.80	100.00
ZER	261885.00	63.98	20.48	6.10	48.90	63.70	80.40	100.00
SIM	262783.00	64.04	21.95	4.40	47.90	62.70	82.00	99.50
SIO	262944.00	67.75	19.43	8.60	52.70	70.90	84.30	100.00
VIS	262033.00	67.85	21.10	8.60	50.60	71.40	86.30	100.00
MVE	262944.00	68.21	20.11	11.60	52.50	68.60	86.20	100.00
BIN	262638.00	68.97	20.69	8.10	53.00	69.60	87.20	99.30
EVI	262621.00	70.34	19.09	11.10	56.00	71.40	86.80	100.00
MTE	262944.00	71.21	19.29	11.10	57.00	73.50	88.50	99.20
BLA	262944.00	72.56	20.43	6.80	59.00	75.60	90.20	99.30
MAR	262891.00	72.68	19.25	10.70	57.70	73.40	90.70	100.00
ULR	262944.00	75.61	20.27	10.40	60.60	83.90	91.90	100.00
MOB	262944.00	77.98	20.40	10.10	64.70	85.80	94.50	100.00

	Date	Station	Pressure (hPa)
	2023-10-20	EVI	976.7
	2023-10-20	SIO	979.4
	2020-12-28	EVO	979.4
	2020-12-28	ZER	979.9
	2020-12-28	MTE	980.1
	2020-12-28	GRC	980.3
	2020-12-28	BLA	980.7
	2020-12-28	SIM	980.7
	2020-12-28	MVE	980.8
	2020-12-28	ULR	980.9

Table 16: Minimum sea-level pressure values recorded per station

- [47] S. MASOUDNIA AND R. EBRAHIMPOUR, *Mixture of experts: a literature survey*, Artificial Intelligence Review, 42 (2014), pp. 275–293.
- [48] METEOSWISS, *Icon forecasting systems*, 2024, <https://www.meteoswiss.admin.ch/weather/warning-and-forecasting-systems/icon-forecasting-systems.html> (accessed 2025-04-14).
- [49] METEOSWISS FEDERAL OFFICE OF METEOROLOGY AND CLIMATOLOGY, *Open data*. <https://www.meteoswiss.admin.ch/services-and-publications/service/open-data.html>, 2025. Accessed: 29 May 2025.
- [50] L. MITAS AND H. MITASOVA, *Spatial interpolation*, Geographical information systems: principles, techniques, management and applications, 1 (1999), pp. 481–492.
- [51] M. J. MONTEIRO, F. T. COUTO, M. BERNARDINO, R. M. CARDOSO, D. CARVALHO, J. P. MARTINS,

Table 17: Total number of 10-minute intervals in which humidity reached 100%

Station	Humidity
EVI	13
SIO	384
MOB	879
ZER	971
VIS	1267
ULR	1963
MVE	2803
EVO	3880
MAR	5114
GRC	8965

Table 18: Performance metrics for hourly aggregation models.

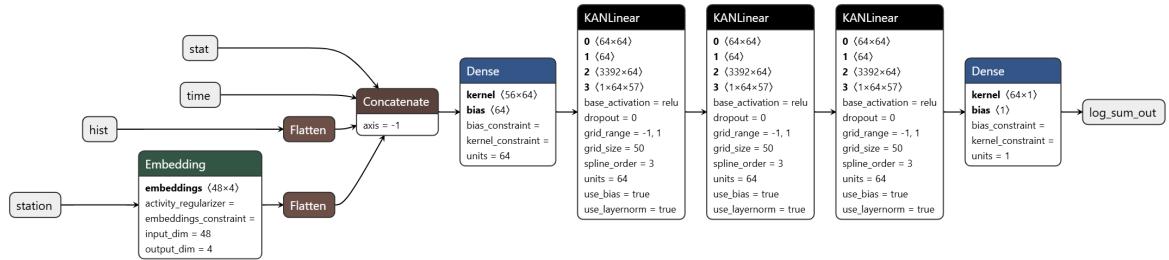
Metric	Score
Precision	0.669
Accuracy	0.823
Sensitivity (Recall)	0.618
F1 Score	0.642
AUC	0.845

(a) Binary classification model.

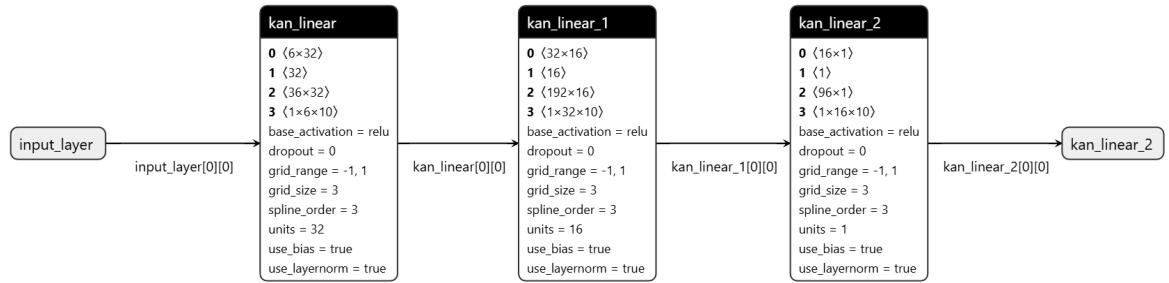
Metric	Score
RMSE	0.579
MSE	0.335
MAE	0.200
R^2	0.005

(b) One-step ahead forecasting model.

- J. A. SANTOS, J. L. ARGAIN, AND R. SALGADO, *A review on the current status of numerical weather prediction in portugal 2021: surface-atmosphere interactions*, *Atmosphere*, 13 (2022), p. 1356.
- [52] T. MORGAN-WALL, *rayshader: Create Maps and Visualize Data in 2D and 3D*, 2025, <https://www.rayshader.com>. R package version 0.38.11, <https://github.com/tylermorganwall/rayshader>.
- [53] S. MÜLLER, L. SCHÜLER, A. ZECH, AND F. HESSE, *Gstools v1. 3: a toolbox for geostatistical modelling in python*, *Geoscientific Model Development*, 15 (2022), pp. 3161–3182.
- [54] NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, *A warming earth is also a wetter earth*. <https://www.ncei.noaa.gov/news/warming-earth-also-wetter-earth>, 2020. Accessed: 19 May 2025.
- [55] S. OCHOA-RODRIGUEZ, L.-P. WANG, P. WILLEMS, AND C. ONOF, *A review of radar-rain gauge data merging methods and their potential for urban hydrological applications*, *Water Resources Research*, 55 (2019), pp. 6356–6391.
- [56] M. A. OLIVER AND R. WEBSTER, *Kriging: a method of interpolation for geographical information systems*, *International Journal of Geographical Information System*, 4 (1990), pp. 313–332.
- [57] S. PANIGRAHI AND H. S. BEHERA, *A hybrid ets-ann model for time series forecasting*, *Engineering applications of artificial intelligence*, 66 (2017), pp. 49–59.
- [58] D. PERLER AND O. MARCHAND, *A study in weather model output postprocessing: using the boosting method for thunderstorm detection*, *Weather and forecasting*, 24 (2009), pp. 211–222.
- [59] K. PHILIPPOPOULOS AND D. DELIGIORGI, *Application of artificial neural networks for the spatial estimation of wind speed in a coastal region with complex topography*, *Renewable Energy*, 38 (2012),



(a) Architecture of logistic regression utilizing only KAN layers. Reaching an AUC of only 0.501, this model is no better than a random decision model.



(b) Architecture of the one-step forecasting model utilizing only KAN layers. This model produced only near-zero forecasts; better data balancing only led to non-converging results.

Figure 31: Model architectures using only Kolmogorov–Arnold Network (KAN) layers. Despite their flexibility, such models fail to generalize effectively.

pp. 75–82.

- [60] I. PRICE, A. SANCHEZ-GONZALEZ, F. ALET, T. R. ANDERSSON, A. EL-KADI, D. MASTERS, T. EWALDS, J. STOTT, S. MOHAMED, P. BATTAGLIA, ET AL., *Probabilistic weather forecasting with machine learning*, Nature, 637 (2025), pp. 84–90.
- [61] S. RAVURI, K. LENC, M. WILLSON, D. KANGIN, R. LAM, P. MIROWSKI, M. FITZSIMONS, M. ATHANASIADOU, S. KASHEM, S. MADGE, ET AL., *Skilful precipitation nowcasting using deep generative models of radar*, Nature, 597 (2021), pp. 672–677.
- [62] S. ROOF AND C. CALLAGAN, *The climate of death valley, california*, Bulletin of the American Meteorological Society, 84 (2003), pp. 1725–1740.
- [63] D. J. ROSILLON, A. JAGO, J. P. HUART, P. BOGAERT, M. JOURNÉE, S. DANDRIFOSSE, AND V. PLAN-CHON, *Near real-time spatial interpolation of hourly air temperature and humidity for agricultural decision support systems*, Computers and Electronics in Agriculture, 223 (2024), p. 109093.
- [64] M. G. SCHULTZ, C. BETANCOURT, B. GONG, F. KLEINERT, M. LANGGUTH, L. H. LEUFEN, A. MOZAF-FARI, AND S. STADTLER, *Can deep learning beat numerical weather prediction?*, Philosophical Transactions of the Royal Society A, 379 (2021), p. 20200097.
- [65] P. SEN, *Weather and weather forecasting*, 2017, <https://api.semanticscholar.org/CorpusID:249010327>.
- [66] I. SIDERIS, M. GABELLA, M. SASSI, AND U. GERMANN, *The combiprecip experience: development and operation of a real-time radar-raingauge combination scheme in switzerland*, in 2014 international weather radar and hydrology symposium, 2014, pp. 1–10.

Hour	MAE	RMSE
1	0.014	0.022
2	0.016	0.025
3	0.017	0.029
4	0.020	0.036
5	0.020	0.038
6	0.039	0.055
7	0.025	0.039
8	0.020	0.031
9	0.019	0.031
10	0.016	0.027
11	0.019	0.030
12	0.016	0.030
13	0.016	0.027
14	0.012	0.026
15	0.010	0.030
16	0.011	0.028
17	0.015	0.048
18	0.049	0.108
19	0.038	0.083
20	0.045	0.078
21	0.025	0.049
22	0.016	0.039
23	0.014	0.036
24	0.018	0.055

Table 19: Hourly aggregated MAE and RMSE for increasing prediction horizon, due to space concerns, only the first 24 hours are displayed.

- [67] S. SOMVANSHI, S. A. JAVED, M. M. ISLAM, D. PANDIT, AND S. DAS, *A survey on kolmogorov-arnold network*, arXiv preprint arXiv:2411.06078, (2024).
- [68] D. A. SPRECHER AND S. DRAGHICI, *Space-filling curves and kolmogorov superposition-based neural networks*, Neural Networks, 15 (2002), pp. 57–67.
- [69] C. TOBIN, L. NICOTINA, M. B. PARLANGE, A. BERNE, AND A. RINALDO, *Improved interpolation of meteorological forcings for hydrologic applications in a swiss alpine region*, Journal of Hydrology, 401 (2011), pp. 77–89.
- [70] H. TORELLÓ-SENTELLES, F. MARRA, M. KOUKOULA, G. VILLARINI, AND N. PELEG, *Intensification and changing spatial extent of heavy rainfall in urban areas*, Earth’s Future, 12 (2024), p. e2024EF004505.
- [71] A. TUEL, B. SCHAEFLI, J. ZSCHEISCHLER, AND O. MARTIUS, *On the links between sub-seasonal clustering of extreme precipitation and high discharge in switzerland and europe*, Hydrology and Earth System Sciences, 26 (2022), pp. 2649–2669.
- [72] C. J. VACA-RUBIO, L. BLANCO, R. PEREIRA, AND M. CAUS, *Kolmogorov-arnold networks (kans) for time series analysis*, arXiv preprint arXiv:2405.08790, (2024).
- [73] A. WIGET, B. VOGEL, AND D. SCHNEIDER, *Die neue landesvermessung lv95 der schweiz*, Zeitschrift für satellitengestützte Positionierung, Navigation und Kommunikation, 3 (1994), pp. 13–20.

- [74] W. M. O. (WMO), *Guide to instruments and methods of observation: Volume iv – space-based observations*, Accessed online: <https://library.wmo.int/idurl/4/68662> (26 May 2025), (2021).
- [75] R. C. YU, S. WU, AND J. GUI, *Residual kolmogorov-arnold network for enhanced deep learning*, arXiv preprint arXiv:2410.05500, (2024).
- [76] N. H. ZAKARIA, S. SALLEH, A. ASMAT, A. CHAN, N. A. ISA, N. A. HAZALI, AND M. ISLAM, *Analysis of wind speed, humidity and temperature: variability and trend in 2017*, in IOP Conference Series: Earth and Environmental Science, vol. 489, IOP Publishing, 2020, p. 012013.
- [77] T. ZUMBRUNNEN, H. BUGMANN, M. CONEDERA, AND M. BÜRG, *Linking forest fire regimes and climate—a historical analysis in a dry inner alpine valley*, Ecosystems, 12 (2009), pp. 73–86.