**INTRODUCTION**

Predicting rainfall is a major component and is essential for applications that surround water resource planning and management. Over the years numerous attempts have been made at capturing rainfall. One area where it is vital to predict the rainfall amount accurately is within rainfall derivatives. Rainfall derivatives fall under the umbrella concept of weather derivatives, which are similar to regular derivatives deﬁned as contracts between two or more parties, whose value is dependent upon the underlying asset. In the case of weather derivatives, the underlying asset is a weather type, such as temperature or rainfall. The main difference between normal derivatives and weather derivatives is that weather is not tradeable. Hence, typical methods that exist in the literature for other derivatives are not suitable for weather derivatives. In this problem domain the underlying asset is the ac-cumulated rainfall over a given period, which is why it is crucial to predict rainfall as accurately as possible to reduce potential mispricing. Contracts based on the rainfall index are decisive for farmers and other users whose income is directly or indirectly affected by the rain. A lack or too much rainfall is capable of destroying a farmer’s crops and hence their income. Thus, rainfall derivatives are a method for reducing the risk posed by adverse or uncertain whether circumstances. Moreover, they are a better alternative than insurance, because it can be hard to prove that the rainfall has had an impact unless it is destructive, such as severe ﬂoods or drought. Similar contracts exist for other weather variables, such as temperature and wind. Within the literature rainfall derivatives is split into two main parts. Firstly, predicting the level of rainfall over a speciﬁed time and secondly, pricing the derivatives based on different contract periods/length. The latter has its own unique problem, as rainfall derivatives constitutes an incomplete mar-ket1. This means the standard option pricing models such as the Black-Scholes model are incapable of pricing rainfall derivatives, because of the violation of the assumptions of the model; namely no arbitrage pricing. Thus, a new pricing framework needs to be established. This paper focuses on the ﬁrst aspect of predicting the level of rainfall. Note that it is essential to have a model that can accurately predict the level of rainfall, before pricing derivatives, because the contracts are priced on the predicted accumulated rainfall over a period of time. In order to predict the level of rainfall for rainfall derivatives, the statistical approach of Markov-chain extended with rainfall prediction (MCRP) is used. Other methods do exist, but this approach in particular is the most commonly used, and will thus be acting as a benchmark for our proposed method-ology. The use of these models allows for the simulation of rainfall on a daily time scale, thus giving more ﬂexibility in the problem domain. The reason why we are interested in daily amounts, rather than monthly or annual amount models is because the models are a lot more ﬂexible to changes. Moreover, one is able to capture trends and more information from studying daily values. Thus, increasing the accuracy of pricing, which is crucial because contracts are priced ahead of time—sometimes this can be up to a year ahead. It is outside the scope of this paper to cover rainfall derivatives in detail. However, the path chosen reﬂects the literature surrounding this application such as and The amount of literature surrounding rainfall derivatives is quite light, due to rainfall derivatives being quite a new concept and rainfall being very difﬁcult to accurately measure. As already mentioned, the use of MCRP is the most prevalent approach, due to its simplicity. The general approach of MCRP is often referred to as a ‘chain-dependent process’, which splits the model into capturing ﬁrst the occurrence pattern, and then the rainfall intensities. The occurrence pattern is produced by calculating the probability of what the out come of today will be given what happened in the previous day(s).The process of deciding upon what state to be in is performed by a Markov-chain, where state 0 is a dry day and state 1 is a wet day. On the other hand, the intensities are produced by generating random numbers from a distribution that ﬁts the daily data. This step is only calculated if we are in state 1, i.e.a wet day. Typically in the literature, the Gamma and Mixed-Exponential distributions provide the best ﬁt for rain data and are most commonly used. We refer the reader to for a complete description of the MCRP approach. However, even though the MCRP approach is quite popular, it faces several drawbacks. First of all, the model is very simplistic and is heavily reliant on past information being reﬂective of the future. Additionally, the predicted amount is essentially the average level of rainfall observed across the study period and does not take into account annual deviations in weather patterns. Furthermore, the model for each city needs to be speciﬁcally tuned as each exhibits different statistical properties, i.e. a new model for each city. Lastly, MCR Produces weak predictive models, as its only focus is on ﬁtting the historical data. This last point is very important, as one should not only be interested in deriving models that describe past data effectively, as it currently happens; instead, we should also be focusing on producing effective predictive models, which can offer us insights on future weather trends. Due to the disadvantages highlighted above, we divert away from the use of statistical approaches and in this paper we propose using a machine learning technique called Genetic Programming (GP). Rainfall prediction has not been covered in great detail within the machine learning literature and the applications are mainly focused on the short term predictions i.e. up to a few hours. Little literature exists for the daily predictions, e.g. used a feed-forward back-propagation neural network for rainfall prediction in Sri Lanka, which was inspired by the chain-dependent approach from statistics. To the best of our knowledge, the only work that exists for daily predictions using Genetic Programming is. However, the GP performed poorly by itself, although when assisted by wavelets the predictive accuracy did improve. However, there has been no previous work in using GP in the context of rainfall weather derivatives. The goal of this paper is thus to explore whether GP is able to outperform the usual approach adopted within the rainfall derivative literature, namely MCRP. GP is chosen for this paper over other machine learning techniques, because it has the beneﬁt of producing white box (interpretable, as opposed to black box) models, which allows us to probe the models produced. Moreover, we can capture nonlinear patterns in data without any assumptions regarding the data. This should allow us to produce a model that can reﬂect the ever changing process of rainfall. As a result, we could capture yearly deviations that the current MCRP is unable to replicate. Additionally, we are able to produce a more general model, which can be applied to a range of cities/climates, without having to build a new model each time. Hence, the main contribution of this paper is that we propose a new GP for the problem of rainfall prediction, and compare its predictive performance against the performance of the current state-of-the-art MCRP approach. This will be the ﬁrst step towards pricing rainfall derivatives using GP.