London School of Economics and Political Science,

Statistics Department,

Centre for the Analysis of Time Series.

Simulation of temperature time-series on long time scales with application to pricing weather derivatives.

Anna Andrianova

A thesis submitted to the Department of Statistics of the London School of Economics for the degree of Doctor of Philosophy, London, October 2008.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

Abstract

Long term weather forecasts are in great demand across many industries, such as the agricultural, tourism, and energy sectors. And hence, one of the major goals of this thesis is to develop a good quality benchmark for long term weather forecasts. In 2001, a survey was conducted on behalf of the UK government by PriceWaterHouseCoopers, which revealed that 95% of UK businesses lost on average 10% from their profits due to unforseen weather. Another goal of this thesis is to develop enhanced methodology that allows business to reduce their weather risk exposure.

There are existing methods and tools that are available to address some of the difficulties of long-term weather forecasting and weather risk hedging. In particular we consider historical weather data and its statistics, synthetic weather data generated by statistical models and short, medium and seasonal physical weather forecasts. For the purpose of reduction of weather risk exposure, weather derivatives are considered. There are however limitations and weaknesses in the currently available techniques.

The quality and availability of historical data varies dramatically depending on the type of the weather variable and the location of interest. Operational errors are not always clearly identified, and are quite often hard to detect. In addition, observational errors are always present. The quality and quantity of the available historical data affects all other methods of weather forecasting. Short term and medium range physical weather forecasts exhibit high skill, but do not cover the desired time range. Seasonal (long term forecasts) are vague in their nature, offering only a probability

of being above, at or below the climate norm, and the notion of the norm is itself questionable. Classical synthetic temperature generation models make questionable assumptions, such as independence of certain observed patterns in the data, and other distributional assumptions. Finally the inapplicability of pricing by dynamic hedging to weather derivatives results in the lack of a unique no-arbitrage price.

In this thesis a new long-term temperature forecasting benchmark is proposed. In particular, the Ensemble Random Analog Prediction (ERAP) dynamic resampler is developed. ERAP allows one to generate temperature scenarios over long time scales without making assumptions about the underlying model of temperature. ERAP works by identifying similar patterns in the historical data across multiple time scales. We also propose a new non-linear weather resembling test system - a weather-like process that mimics the real temperature with additional long term non-linear patterns. Finally we study the mixing of physical weather forecasts with the historical data. In particular, combination forecasts are developed that mix information from both physical forecasting models and historical data. The methodology is developed by exploring kernel dressing of forecast scenarios and ignorance skill-score based optimization of parameters.

The new weather generator ERAP is then extensively tested in the perfect model scenario, by studying its performance in terms of the generated statistics, using both a noisy Lorenz system and the new weather like test process. ERAP is also tested on real weather data by assessing its performance on the Berlin daily maximum temperature in terms of the generated

statistics. Finally ERAP is also used for pricing a weather derivative, and the prices compared to other existing techniques including pricing based on the historical statistics, Monte-Carlo using a fitted distribution, plus other statistical techniques from the weather derivative literature. For combination forecasts we study the sensitivity of parameters to ensemble size and the level of noise in the initial conditions, within the perfect model scenario.

We show that ERAP performed well in the perfect model scenario, on the actual data and when used for pricing. The historical statistics were closely replicated, the statistics of a chosen verification set were also well replicated and in some cases, ERAP generated data provided a better match to the statistics of the verification set than the climatology (the statistics of the learning set itself). Some non-conventional statistics were better replicated using ERAP in both the perfect and imperfect model scenarios. Additionally, information is provided by ERAP on the uncertainty of the computed statistics. We also show that ERAP provides more reliable pricing, because it provides more reliable long-term simulations.

The fake weather generator developed in this thesis has shown to provide a good test data set that is non-linear with patterns on multiple time scales and closely resembles the characteristics of real temperature time series. This process could be a viable alternative, if parameters are fully calibrated to the chosen weather data, to existing statistical temperature modeling approaches.

This work could be further improved by the creating better parameter estimation techniques for ERAP. ERAP could also be extended to several dimensions allowing the generation of more enhanced synthetic weather data. For the purpose of pricing weather derivatives more work needs to be done to address the transformation of ERAP scenarios to probabilistic weather forecasts. Further work may also include studies of the performance of combined forecasts, which mix synthetic data and physical weather forecasts, in practice.

Acknowledgement

I would like to dedicate my thesis to my father and all the people battling with cancer. I hope this work will make him feel proud. It has been a long journey with endless hurdles but i am glad that I saw it through till the end. It made me the person I am today.

I want to give special thank you to my supervisor Lenny Smith for all his support and guidance. For not giving up on me through all these years and for opening my eyes to so many ideas and concepts. It is fun to rediscover the universe in the different light.

I want to give endless gratitude to my mother for all her support, for all her advice and love. For always being there for me and for giving me the strength and courage to go on. For always picking me up like no other person ever could. For raising me and being my best friend.

I also want to thank David for all the late nights debating maths, for his patience and love towards me. Most importantly for holding me in my worst moments. This thesis would have not been finished if i was not lucky enough to have such a wonderful person next to me. He has been my inspiration.

I want also to mention my dearest fellow PhD student Milena. I cant believe our late nights in the office will never happen again. Thank you so much for being a great friend and all our coffees together. I will treasure endlessly our times crying on each other shoulders and talking chaos.

Last, but not least, i want to thank all my friends, work colleges and family members for their patience and understanding and for always believing in me.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction.

The aim of this thesis is to build a methodology for weather index modelling, which will allow us to improve the valuation of weather contracts in all possible pricing frameworks. Additionally we aim to provide a high-quality long term benchmark forecast for weather predictions derived from physical models.

Weather forecasting has been an area of great interest and great development. Even hundreds of years ago people were trying to forecast the weather by noticing certain recurrences, for example, that a cloudy sky at night might precede rain and milder temperatures. Although weather forecasting skill has improved significantly since, many questions still remain unanswered.

Over the years weather forecasting became rather sophisticated and at present there are different time-scale forecasts that are available to a user (discussed in chapter 2) including seasonal forecasts, which extend a relatively long time into the future. Seasonal forecasts, however, are extremely limited in their use due to their nature. In this thesis we address this issue of long term forecasting, in particular in chapter 4 we develop a technique for combining information from climatology and physical forecasts.

Apart from forecasts physical, limited historical data and synthetic weather data are also available. This allows us to compute statistics that are often

used as a forecast, for longer time scales and as a benchmark for shorter range forecasting. In chapter 5 a dynamic climate generator is developed that produces enhanced long-term climatology- based forecasts and offers an improved benchmark for the short range forecasts. The generator is extensively tested on data sets constructed in the perfect model scenario and then applied to real temperature data. The construction of data sets for the generator experiment and a description of the real temperature data are provided in chapter 3.

A wide variety of industries are exposed to weather related risks. In August 2001 an independent study was conducted on behalf of the Met. Office, [**9**] on the impact of weather on British businesses. The survey was carried out on businesses from a wide variety of areas such as: leisure, retail, manufacturing and finance. The survey revealed that 95 percent of the UK firms questioned have lost 10 percent of their profits due to unexpected weather. A new opportunity has arisen, in the form of weather derivatives, that allow businesses to offset at least some of their weather exposure by transferring risk into the financial markets.

The weather derivatives market was born as a result of deregulation in the power market, when energy traders Aquila, Enron and Koch Industries constructed and executed the first weather derivative contract [**15**]. Since 1997 the market has been growing; the number of contracts traded has increased as has their variety ([**9**] and [**14**]). Although the weather derivative market has been growing, there are two main factors that slow its development: deficiency of historical weather data, discussed in chapter 2 and the

lack of a systematic pricing framework. The latter problem is mainly due to the difficulties in weather index modelling, which is inherited from the limitations in weather forecasting.

We combine the techniques developed in this thesis in order to construct a coherent and improved framework for modelling of weather indices (chapter 6) and as the result an improved weather derivative pricing methodology is produced. In particular, a weather derivative is priced using a new dynamic climate generator in an actuarial pricing framework. This is followed by a discussion on the incorporation of combined long term forecasts into the pricing method. Finally, in chapter 6 conclusions are drawn and future work is proposed.

CHAPTER 2

Background: Weather and Weather Derivatives.

In this thesis a new hedging approach for weather sensitive risk is proposed. Successful weather risk hedging requires: information on future weather behaviour; and the formulation of a hedging strategy that would allow us to compensate for the weather risk exposure. Prior to presenting the new proposed weather hedging approach, in this chapter we examine what has been developed in recent years in both: the area of weather forecasting/modelling; and the area of weather hedging.

In particular, in the first part of the chapter, three different categories of techniques that are used to describe and model weather behavior are discussed. These three categories are: climatology, synthetic weather data and physical weather forecasts[1]. First, issues concerning historical weather data and climatology based forecasts are examined (section 1.1). Followed by a discussion on modelling approaches that are used to construct synthetic weather data (section 1.2). Finally, the characteristics and limitations of physical forecasts that fall into three categories: seasonal, medium range and short term, are considered (section 1.3).

The second part of the chapter is an overview of the development of an extremely versatile weather hedging product - the weather derivative. First a brief description of the market is presented, followed by the characteristics

[1]For the definitions please refer to the Glossary for this chapter.

of weather derivative contracts. Finally the approaches currently used in pricing weather derivatives are listed (section 2).

1. Weather.

As previously stated, accurate identification of the future behaviour of the weather is a major step towards successful weather hedging. Approaches that are currently used in describing and modelling weather behaviour can be separated into three main categories. These categories are: climatology, synthetic weather data and physical weather forecasts. The three categories vary in their use, mainly depending on the resources available and the time scale for which the forecast is required.

The first category - climatology, obtains information about weather behaviour using historical weather data only. In particular, information obtained does not involve any modelling, simply various statistics calculated directly from the historical data. In contrast, the synthetic weather approach requires a statistical model for the behaviour of weather data. This approach does not model the physics of the weather system, but rather fits a statistical model to a chosen weather time series (such as temperature or the precipitation data). As part of the new weather hedging approach (chapter 5) we propose a new synthetic weather generator that provides improved weather risk hedging across longer time scales. Finally, the physical weather forecasting models describe the evolution of the weather system according to physical laws. Such forecasts are usually produced by meteorological institutions, such as Met Office.

Although, each of the above approaches have weaknesses and limitations (discussed in sections 1.3, 1.1 and 1.2), they are capable, especially when combined, of producing a useful description of weather behaviour that can be used to model weather sensitive risk exposure. In chapter 4 we propose a new approach to combining information from physical forecasts and synthetic weather data that can be used in pricing weather derivatives (see chapter 6 on the discussion of further application of this method).

1.1. Historical weather data and climatology. *Climatology* is defined to be historical statistics of weather data such as historical frequencies and averages of observations [**3**].

Collecting descriptive statistics of the historical weather data is a crucial step towards modelling and or predicting weather behaviour. It can either be used on its own as a weather modelling tool or can be used as a first step in simulated weather approaches. It also serves as a benchmark for all other weather forecasting techniques.

Weather data is used not only in climatology but in all other areas of weather modelling. Although, there has been a significant improvement in the quality and availability of weather data, there are still several issues affecting the data, and as a result the statistic obtained from it:

- The weather data record is rather limited. Temperature records extend further than any other weather variable, and even so the time span it covers is minute compared to the existence of the global weather dynamics. The lack of historical data has been a

significant motivation behind the development of synthetic weather generators.

- The available data quite often contains missing values that are sometimes not clearly indicated or have been inappropriately substituted. This is another reason why synthetic weather generators have been developed.
- Before satellites, the number of locations on the globe where data has been recorded, compared to the total surface area, is very small. At some sites data was not recorded due to the economic, social or political circumstances. Historical measurements are particularly course in oceanic regions [4]. This is one of the sources of errors in physical weather forecasting.
- Measurement error and human error, that are inevitably present in data, are very important factors that are difficult to identify. Physical weather forecasts are particularly sensitive to measurement error. Measurement error has been one of the reasons why physical weather forecasts are now produced in the ensemble mode[2].
- Finally, even if the data is 'perfect', statistics vary greatly depending on the amount of data used in the analysis, due to the observed recent trends in temperature and other variables [5]. This has caused a deviation in opinion among practitioners in the financial markets on the amount of data that should be used in modelling weather risk hedging strategies. The new proposed synthetic

[2]See glossary of this chapter for the definition and section 1.3 for the explanation on ensemble forecasts.

weather generator, chapter 5, allows us to avoid the debate on the size of historical weather record that is used in modelling weather risk.

1.2. Synthetic weather data. Producing synthetically generated weather data is a widely used weather forecasting approach, specifically in the area of managing weather sensitive risk. Typically the synthetically generated data is used whenever physical weather forecasts are not available. Weather generators produce synthetic weather data, such that it is consistent with the historical data, often offering enhanced statistics. Such forecasts also can be used as benchmark forecasts. Weather generators are statistical models that can fill up missing data or produce indefinitely long synthetic weather time series. Usually key properties of observed meteorological records, such as daily means, variances, extremes, etc are simulated.

Most techniques developed use a stochastic framework (stochastic weather generators) to model a particular weather variable. This allows a certain flexibility in order to maintain uncertainty due to imperfect modelling[3]. Some weather generators were focused on modelling precipitation. The motivation for modelling precipitation was driven by the end users, many of whom required precipitation forecasts on a long time scale. Additionally, the occurrence and amount of precipitation affects the statistics of other weather variables, such as temperature(Wilks and Wilby, 1999). As the result, a new class of weather models was developed. These models first model precipitation occurrence and intensity, and then condition the statistics of

[3]For the definition on imperfect modelling please refer to the glossary of this chapter.

daily non-precipitation variables on the occurrence or non-occurrence and amount of precipitation [**28**].

Most of the statistically based synthetic weather models used, however, generate daily temperature data directly. In this thesis we concentrate on the models of synthetic temperature, that have been proposed for the purpose of managing weather risk. Models in this category typically assign a functional form to different behavioural patterns that are generally observed in temperature data. Such patterns will include, but are not limited to: seasonality, daily oscillations, past state dependance and long term trends. Alternatively, models in this category may make assumptions about the distribution of temperature, which can then be used directly or re-sampled to produce synthetic temperature scenarios.

Practitioners, particularly in the financial sector, are often more concerned with replicating the way that other practitioners in the market model their risk exposure, rather than modelling future weather more precisely. For that reason we pay particular attention to the most used synthetic weather models (in the financial sector) and use the output from those models, generated in chapter 3, as the benchmark for the new proposed weather generator. In particular, the methodologies proposed by Cao and Wei [**30**] and M. Davis [**31**] are examined in chapter 3 and are used to generate synthetic maximum daily temperature for Berlin, Germany. The comparison is then made by pricing a weather derivative using those approaches and the new proposed synthetic weather generator(chapter 5 in chapter 6).

1.3. Physical weather forecasts. Physical weather forecasts are based on the meteorological weather modelling approach, where the physical dynamics of earth's atmosphere is considered. There are three main categories that physical forecasts fall into, depending on their time scale: short term, medium range and seasonal forecasts. For short and medium range forecasts the evolution of the weather system is modelled. For seasonal forecasts, long-term weather subsystems are studied, in particular, the weather patterns that had been observed in the past whenever certain long-term weather subsystems were present [**16**]. This modelling approach offers potentially the most accurate weather forecast especially on short time scales. The equations that describe the evolution of the weather system, however, exhibit sensitivity to initial conditions, such as you see in chaotic systems[4] [**27**]. We proceed by examining the characteristics of each of these categories in more detail.

Short term forecasts usually extend up to four to five days maximum. In recent years short term forecasts are produced in the form of an ensemble forecast Each ensemble member represents a possible temperature scenario at a given time, rather than an exact prediction. This allows us to account for uncertainty due to measurement error in initial weather observations. An ensemble forecast is constructed by using perturbed initial conditions for the equations of motion that model the weather dynamics. This is further extended by producing probabilistic weather forecasts, where each ensemble member, at a given time, is assumed to be a possible draw from a forecast

[4]On the definition of chaotic system please refer to Glossary of this chapter.

distribution [6]. A particular feature of short term forecasts is that ensemble members usually exhibit relatively similar behaviour to each other for up to two or three days, providing that a good model is used to construct such an ensemble [6]. Even on such a short time scale on accurate forecast is not guaranteed. It has been stated by Tim Legg and Ken Mylne, Predictability and Ensemble Forecasting Group, NWP Division, UK Met Office, that: 'forecasts beyond day 3 can sometimes be completely wrong, and even shorter-range predictions can at times contain serious errors, at least in details or timing' [1].

Figure 1 illustrates a one day forward probability forecast for temperature constructed using 51 ensemble members (MetOffice, UK, [2]). Short term forecasts exhibit less uncertainty compared to medium and seasonal range forecasts [7] and are very useful if predictability is required for a short term, particularly in the time period of less than three days. The procedure involved however in producing such short term forecasts is extremely complex and requires significant resources. It is important to note that the distribution produced is not in a standard form (such as Normal Distribution [25] for example), at least in example provided.

Medium range forecasts are very similar in their characteristics to short term forecasts, however the main difference is that forecasts produced are for the longer term, which extends to ten to twelve day time period, where the uncertainty is much greater. Typically as time increases the spread, given by the Euclidian distance ([8]) between ensemble members on a given day, in the ensemble gets larger. As a result, a wider spread in the forecast

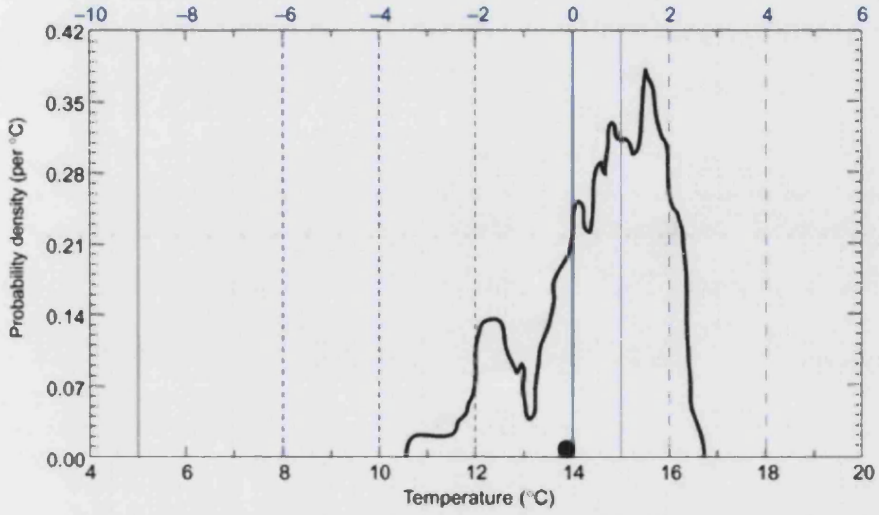


FIGURE 1. Example of one day forward temperature probability density forecast, produced by PREVIN - the numerical weather forecasting system of the UK Met Office. The top scale shows temperature anomalies (w.r.t climatology); the operational forecast value is marked as a circle.

distribution will be observed. On figure 2 an example of a five day forward forecast probability distribution for the temperature at Heathrow airport is presented. It can be noted that the difference between the maximum and the minimum temperatures with positive probabilities is larger compared to the difference observed in the one day forward distribution, figure 1. It is also evident that the distribution is wider (in terms of standard deviation [**25**]) and the shape is more complex exhibiting several peaks.

Medium range probability forecasts offer much greater potential rewards compare to short term forecasts, however at a greater cost. Uncertainty is much larger and the computer resources needed are much greater. Both

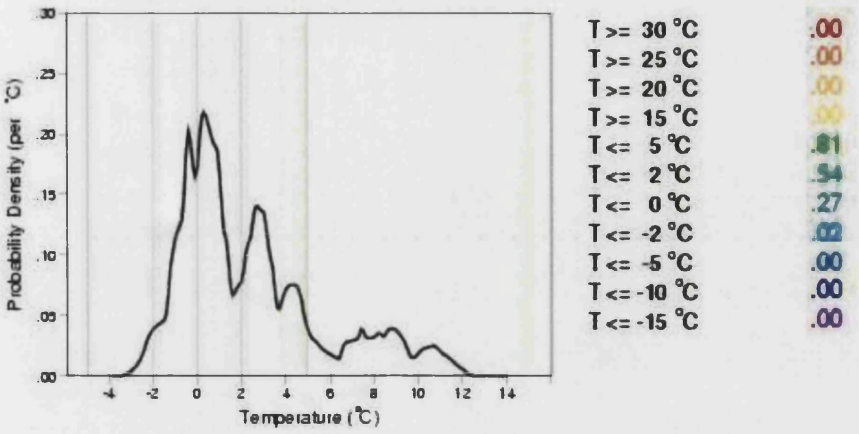


FIGURE 2. An example of a 5-day forecast of the relative probabilities of different temperatures at Heathrow Airport for midday on 28th February 2004, produced by the UK MetOffice.

short term and medium range forecasts will offer little no reward for long term weather forecasting.

In terms of successful weather risk hedging strategy, forecasts are required on a much longer time scale than medium range forecasts can offer. The main characteristic of physical seasonal forecasts is that they are vague in their nature, offering probabilities of a divergence from climatology of a particular weather variable, rather than a day to day fluctuational forecast (for example see [**16**]), see figure 3. This seriously limits the use of seasonal forecasts on their own, specially in the case of weather derivatives.

Producing seasonal weather forecasts involves analysis of large atmospheric and atmospheric-oceanic systems of the global weather dynamics. Forecasting in this framework is possible due to the operation of these systems on a longer time scale and statistical association of these systems with

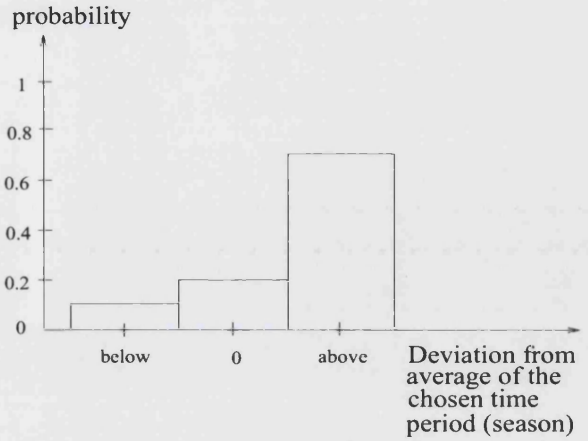


FIGURE 3. An example of a seasonal forecast for a given variable. The forecast issues probabilities for that variable being above average, below average or average.

certain weather patterns [**16**]. The reliability and length of such forecasts vary through regions, as some of the atmospheric-oceanic systems are more predictable in their behaviour and are slower progressing than others.

Figure 4 illustrates short term, medium range and seasonal forecasts, in particular, how each of these relate to each other with regards to the time scale. Medium range and short term forecasts are presented in the ensemble mode.

2. Weather derivatives: an overview

It has been stated that identifying future weather is a crucial step towards successful risk hedging. The other major component involves the identification and modelling of weather risk exposure. A financial instrument, the *Weather Derivative*, was introduced onto the market in 1997 [**15**], that is flexible enough to allow most weather sensitive industries to cover themselves against weather risk.

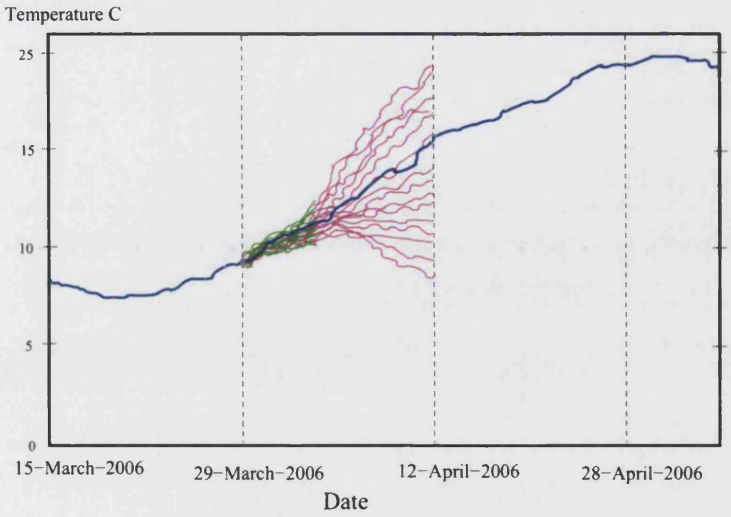


FIGURE 4. Comparative illustration of forecasts from three categories of physical forecasts, characterised by the time scale that they are produced for. Blue- a seasonal forecast, magenta- an ensemble of a medium range forecast and green- an ensemble of the short term forecast.

A Weather Derivative is a financial contract for which the payout depends on an underlying weather variable. If certain conditions of the weather occur a purchaser of a contract receives a payout, which is usually proportional to the change in a weather variable, although in some cases the payout is fixed.

The weather derivative market has two types of participants: speculators and weather sensitive end users, see Figure 5. Weather sensitive end users were the initiators of the market. These businesses want to hedge their exposure to the weather by purchasing weather derivative contracts that would cover, or at least partially cover their existing weather risk. The sellers of weather derivatives are mainly from the financial sector, these companies

are not exposed to the weather risk until they enter into a contract. The profit in this case would heavily depend upon reliable hedging against any possible losses.

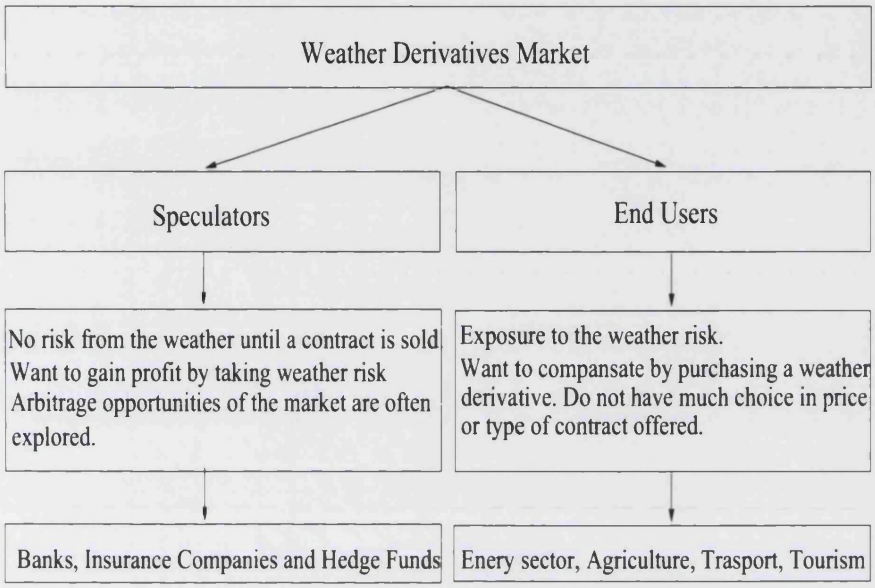


FIGURE 5. Diagram of the participants of the Weather Derivative market and their characteristics

Initially there was a fast development of the weather market, as it offered a sophisticated weather hedging strategy and additionally an uncorrelated product to many financial instruments on the market [**9**]. At present however, the weather derivatives market is extremely illiquid[5], with a very small number of trades, and lopsided, where most participants are potential buyers of contracts, which pushes up the prices of weather derivatives, hence further reducing liquidity. The main reason behind this is the lack of a standard weather pricing methodology combined with the uncertainty in modelling the weather (temperature in particular) long term. The new

[5]See glossary for definition

weather hedging approach proposed in this thesis uses weather derivatives as a hedge structure, but evaluated using on improved long term weather modelling framework. In chapter 6 a weather derivative is priced using different weather modelling approaches, including the new proposed synthetic weather generator. For that reason, let us consider the structure of a typical weather contract.

In a typical weather contract, the following is specified:

- The length of the contract, which might range from one month to up to (typically) a season, and its starting date, which we will refer to as t_0. T - time when the contract expires is usually referred to as the expiry.
- Underlying Index, will always be a weather variable or a function of weather variables. In this thesis we price a weather option that is written on a Cooling Degree Days (CDD) index, figure 6. Degree Days are defined to be: number of degrees, above (for the CDD) or below (for the Heating Degree Days) a chosen threshold[6] added over the length of the contract $T - t$ days. The arithmetic average A_i of the observed daily maximum and minimum temperatures is calculated in order to determine by how many degrees the temperature on a particular day i exceeds or goes below the threshold, equation (1).

[6]in Europe this threshold is defined to be 18 degrees Celsius

(1) $$CDD = \sum_{i=1}^{T-t} \max(0, A_i - 18)$$

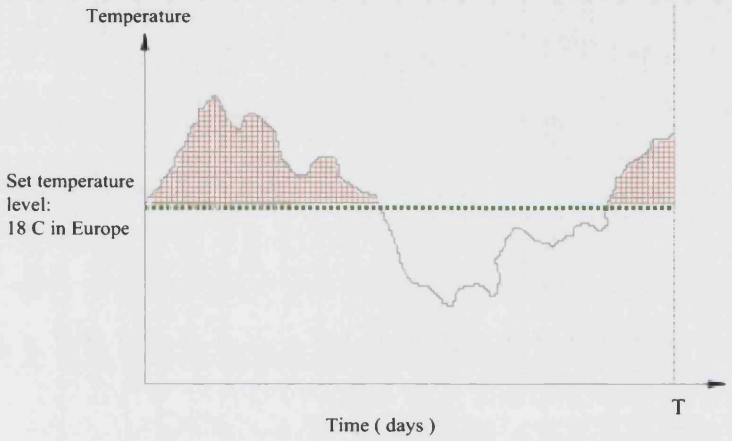


FIGURE 6. Example of the cooling degree days index calculation over the length of a contract, where the dotted green line indicates the set level of the threshold and the red crossed area indicates the days which will qualify and as the result the temperatures that will be used in the calculation of the CDD index.

- Location: the locations for which weather contracts are structured at the moment are limited. In theory any location could be chosen, provided there is weather data available.
- Type of contract: this section describes the rights given to a buyer of the derivative. Some standard contracts, the payoff functions of which are illustrated in figure 7, include: a *Call Option* that gives a purchaser the right to buy a specified asset at the agreed strike price at a specified time in the future and a *Put Option* that gives a

purchaser the right to sell a specified asset at the agreed strike price at the specified time in the future ([**24**]). For weather derivative the asset is cash proportional to CDD. The weather derivative priced in chapter 6 is a call option.

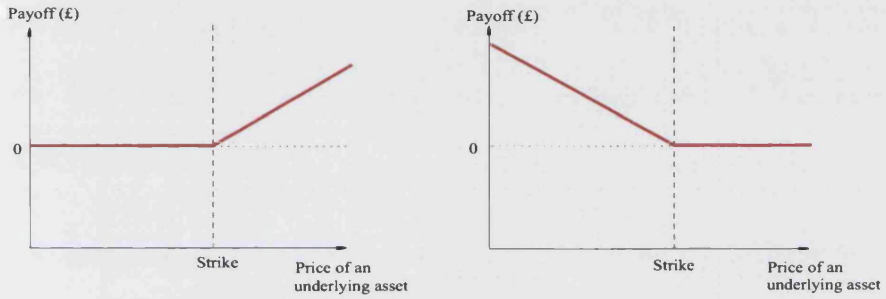


FIGURE 7. Payoff function of a Call Option (right corner), where the payoff is zero until the underlying asset reaches the value of the predetermined strike. If the value of the underlying asset is bigger than the strike the payoff is determined by the difference between the two. Payoff function of a Put Option (left corner), where the payoff is zero unless the asset's value is less than the value of the strike. Although in this figure the payoff (in both cases) does not go negative, in practice if the price of the underlying asset is below (for call) and above (for put) the strike price, the payoff is at the negative level of the option premia.

- Choose tick value, or fixed payout amount. Tick value attaches monitory value to the underlying weather variable index. Usually, to a degree, say, there would be an amount attached, which gets multiplied by the index in the end of the contract and with the

help of the pay off function the payout on the claim is calculated in pounds, say, rather than degrees Celsius.

2.1. Pricing Weather Derivatives. Whenever a derivative is priced, a discounted expected payoff under a pricing measure is determined [**17**], equation 2.

$$V_{weather} = exp\left(-r\left(T-t\right)\right) E_M\left(P\left(I\right)\right), \tag{2}$$

where $P(I)$ is a payoff function that maps the underlying index I to the claim of the contract and M is the pricing measure. $exp\left(-r\left(T-t\right)\right)$ is the discounting factor under the assumption that the interest rate r is continuously compounded.

In the financial markets, values of contracts are based on the Black-Scholes no-arbitrage pricing [**22**] framework. The Black-Scholes method provides a unique pricing measure M, and hence a unique price V, for any given contract, together with a perfect hedging strategy, provided that the conditions of the model are satisfied and the market is complete[7]. One of the crucial conditions required by Black-Scholes is that the underlying asset of the derivative has to be traded. Temperature and other weather variables are not traded assets, and the weather derivatives market is illiquid and hence incomplete, which unfortunately forces one to abandon the well developed no-arbitrage[8] framework of pricing.

[7]For the assumptions of the Black-Scholes model please refer to [**22**] and see glossary for the definition of the complete market

[8]See glossary for definition.

Due to inapplicability of the Black-Scholes approach, in order to price a weather derivative according to equation 2, one has to identify:

(1) the pricing measure M;

(2) how to model/forecast index I.

In this section existing approaches of pricing weather derivatives are examined. The existing methods of pricing fall into two different groups: one of which concentrates on modelling the underlying weather index and other addresses issues concerning the pricing measure. As the emphasis in this thesis will be made on the modelling of the behaviour of temperature based indices, the latter will be visited only briefly in this section.

2.2. Modelling behaviour of the underlying index for the purpose of pricing weather derivatives. In the weather derivative pricing literature an underlying weather index most often modelled using:

- Climatology based approach and/or
- Synthetic weather generator.

2.2.1. *Climatology based pricing approach.* The first method that was proposed for the problem of pricing weather derivatives was based on the climatology of a weather variable, in particular the *Sample Mean* [**25**]. Pricing a weather derivative using the sample mean based on various sets of historical data was described in [**29**] by Zeng (2000) and is still a widely used preliminary step in pricing and hedging weather derivatives. It must be noted that the sampled mean price should be used as a benchmark rather than the absolute quoted price, as it is only a fair price for both parties under many consecutive runs in the repeated experiment. A mark up, often in

the form of sampled standard deviation is added to the sampled mean price in order to hedge some of the risk.

Another entity that is used in determination of the hedging strategy is *Value At Risk (VAR)*, [**24**]. Value At Risk is given by some value, v say, such that the probability that the payout is greater than v is $q\%$, in [**29**] $q = 10\%$. VAR is also computed in chapter 6 using different competing methods. VAR also can be used as a price of the weather derivative, providing that the $q\%$ is not set too high, as this would result in a too expensive price for a buyer. The value of $q\%$ will change, depending on the risk preference of the user, the more risk-averse [9] the user, the higher the $q\%$. The climatology only based pricing is often referred to as the actuarial method of pricing.

The actuarial method could be further extended by fitting a distribution to the historical index, often using calculated sampled mean and standard deviation and assuming Normality of temperature data. The distribution fitting approach was proposed in [**29**] in order to satisfy two purposes. First purpose is achieving a more secure hedge on the option, where instead of using the standard deviation as the mark up on the sampled mean price, a more conservative value could be used based on the tails of the fitted distribution. The second purpose is the desire for larger amounts of data, where additional data is generated using the Monte Carlo technique, [**17**] based upon the fitted distribution.

[9] See glossary for definition

In [29], the expected contract payout, volatility and VAR, calculated using only historical data, are compared to the expected payout, volatility and VAR calculated using long run estimates that are generated via a Monte Carlo simulation approach. Zeng demonstrated that the Monte Carlo approach produced larger estimates for all the variables of interest. Additionally, both historical estimates and Monte Carlo based estimates of the expected payoff, volatility and VAR are compared for different quantities of historical data used in the analysis, in order to find the sensitivity of the result to the amount of data used. The value of the sampled mean and standard deviation would vary dramatically, as the number of data used changes. The values increased as the number of years used in calculation decreased, which is consistent with problems discussed in section 1.1. This gives another advantage to the proposed new synthetic weather generator.

In chapter 6 Zeng's experiment is reproduced, where the weather derivative, used as an example in [29], is priced using: the climatology based approach (with the historical data sets that are consistent with the data sets chosen by Zeng) and the Monte Carlo based approach. We further compare results of climatology only based prices with competing approaches, including: the new proposed weather generator (chapter 5) and other competing synthetic weather generators (that are discussed in chapter 3).

The sample mean method is the easiest to use and implement, however it is not clear how many years of data should be used in the calculation. By fitting a distribution to the historical data one can achieve extra flexibility in the determination of the mark up price and perhaps address, to some

extent, the lack of data, however with added problems. Apart from suffering from the same constraints - the mean and variance sensitivity to the amount of data used in the analysis - a further assumption is introduced by choosing a shape of the distribution that is fitted. The value of the mark up entirely depends on the tails of the assumed distribution. This gives another advantage of using the new proposed synthetic weather generator (chapter 6) that does not make any assumption on the distribution of the underlying weather variable but at the same time produces the desired amount of synthetic data.

2.2.2. *Synthetic weather generators based pricing.* There are several models of temperature data that have been proposed in the weather derivative pricing literature. The pricing framework that is used as a benchmark for the new proposed synthetic weather generator for the purpose of pricing weather risk, was proposed by Cao and Wei [**30**]. Cao and Wei (2004) proposed to model the underlying index, temperature in particular, by assuming that temperature can be modelled as Seasonal Normal Temperature (SNT) [10] with an Autoregressive Process of order three [**26**] (AR(3)) plus and seasonal volatility. For the details of Cao and Wei's model refer to chapter 3, where the methodology of Cao and Wei is reproduced to construct synthetic temperature data for Berlin, Germany. Another, similar modelling approach was proposed by Davis (2000) [**31**], where temperature is also modelled using SNT and an Autoregressive Process of order one. Davis's methodology is also reproduced based on the Berlin data in chapter 3.

[10]Deffinition in glossary

Other authors, such as Campbell, Diebold (2005), suggested to use simple statistical approaches to model the weather time series. Specifically, the authors claim that the generated data have the 'conditional mean and variance dynamics preserved', see [**32**]. Temperature is modelled using a simple low-order polynomial deterministic trend, autoregressive lags and a low-order Fourier series to model seasonality.

Another, very popular temperature modelling framework that was proposed is in the form of a stochastic processes[11] [**23**]. This class of model is less widely used for pricing weather derivatives compared to the simple statistical models. Mean reverting[12] [**18**] models driven by a standard Brownian Motion [**19**], were initially and are still used as a basis for Monte Carlo Algorithms, for example see [**33**].

Modelling temperature time series with a stochastic process requires an assumption about the shape of the distribution. Modelling certain signals in the temperature data separately assumes that those signals can be modelled independently. Many authors have used simple temperature models, despite the realisation that such models are not necessarily the best possibility for the weather data modelling, in order to concentrate and address the second half of the problem of pricing weather derivatives: the pricing measure.

[11]See Glossary for the definition

[12]See Glossary for the definition.

In chapter 6 a weather derivative is priced using Cao and Wei and Davis's methodology, in the actuarial pricing framework. The results are then compared to Zeng's (climatology based) prices and prices obtained using the new proposed synthetic weather generater.

2.2.3. *Pricing weather derivatives using different optimisation mechanisms.* Different pricing measures in the calculation of the expected value (equation 2) result in different prices. For example, if the real measure is used the price obtained is the fair price under many repeated experiments; if the risk-neutral measure [**17**] is used the price is fair if the market is complete. Alternative optimisation approaches to pricing under the real measure, in the incomplete market setting is also a big area of research.

One of the most popular, and mathematically interesting approach to pricing weather derivatives involves maximisation of different *utility functions*[13]. Utility functions allow different risk attitudes to be attached to different investors. In order to find a price for a given derivative, the expected value is taken of the utility function of a payoff. Often when utility functions are used, the closed form solutions for the option prices can be derived, by making assumptions on the behaviour of the underlying asset, [**30**]. Some authors also use the utility approach together with the statistical modelling of the underlying weather variable, see [**30**].

2.2.4. *Pricing weather derivatives using seasonal weather forecasts.* In the final section of this chapter lets examine the use of weather forecasts in weather derivatives pricing. Several authors propose different approaches to

[13]See glossary for the definition of a utility function

incorporating extra information received from weather forecasts. One has to be particularly careful when trying to incorporate weather forecasts that are given for a shorter time scale than the length of the contract.

Briggs and Wilks, [**42**], present a procedure for estimating climatological statistics for a broad range of subseasonal variables, conditional on seasonal forecast probabilities, by bootstrapping the observed climatological record consistent with the forecast probabilities. Their procedure computes conventional climatological statistics using weights equal to the probabilities specified in the forecast, for more details see [**42**]. This idea has been further explored by several authors, who statistically model the probability function of the weather variable and bootstrap the distribution of the weather variable according to the forecasted seasonal probabilities. this distribution is then used in the expected value calculation, [**43**].

The weather derivative in Zeng's paper,[**29**], is also priced using seasonal forecasts and the results obtained are compared to the Actuarial method of pricing. In particular, the long term average, which is referred to as the climate norm is based on historical data from 1961 and 1990. The long term forecasts extend for three months. Specifically, the predictions for June, July, August were given by the National Centre of Environmental Predictions. The assumptions are made that: the seasonal summer weather predictions approximate the July prediction and that the probabilities that the CDD will be above, near and below the climate norm are well approximated by the seasonal forecast probabilities of temperature being above, near and below norm.

A normal distribution is then fitted to the historical data for the calculated index (Cooling Degrees Days), by using sampled mean and sampled variance. The resulting distribution is then divided into three equal probability areas with below, near average and above average index values. Finally, the fitted distribution is sampled in accordance to the given forecasted probabilities being above, below and near average. This approach is sometimes called a biased sampling Monte Carlo approach.

In chapter 4 we propose a new methodology for combining information from short/medium range physical forecasts and historical data. The methodology is developed in a perfect model scenario, and tested on chaotic and Autoregressive type systems. It is then applied to combine information form short/medium range physical weather forecasts and the generated synthetic weather data. This is particularly useful when weather derivatives are revalued daily during the life of the contract[14].

[14]The time between the starting time of the contract t_0 and the expiry time T (refer to section 2 for more detail.)

2.2.5. *Glossary of terminology used in this chapter and definitions.*

Climatology	Statistics of historical weather data.
System	is a set of interacting or interdependent entities forming an integrated whole. In the case of this thesis the system is described in the form of a mathematical function.
Chaotic System	Systems whose state evolves with time that may exhibit dynamics that are highly sensitive to initial conditions. This sensitivity manifests itself as an exponential growth of perturbations in the initial conditions on a trajectory.
Ensemble Forecast	Forecast that contains several possibilities (scenarios) at each forecasted time t.
Imperfect model	Model that approximates the actual system.
Illiquid Market	Market where assets are infrequently traded or unavailable.
Complete Market	is one in which the complete set of possible gambles on future states-of-the-world can be constructed with existing assets.

No-Arbitrage	when there no possible advantage can be taken of a price differential between two or more markets.
Risk-aversion	Risk aversion is a concept where an investor would have a preference for a more certain but lower payoff to a less certain but higher payoff.
Seasonal Normal Temperature	An average daily temperature, where an average is taken for ever day of the year. over multiple historical years. So average 1st of January, average 2nd of January, etc.
Utility function	A functional form for the risk preference of an investor.

CHAPTER 3

Observations, Synthetic weather data and a 'Weather like' test data.

In this thesis we propose a new approach to hedging weather risk in the form of an improved synthetic temperature generator for long time scales applied to pricing weather derivatives. The purpose of this chapter is to generate benchmark synthetic weather data based on alternative methods proposed in the weather derivatives literature. Additionally to generate our own 'weather like' test data set that in its behaviour mimics the components of weather data, such as seasonality, daily oscillations and weather fronts. Finally we also provide descriptive statistics for the actual weather data that is used through this thesis.

In particular, the historical record of Berlin daily observed maximum temperature is analysed to produce descriptive statistics section 1. This is followed, in section 2, by implementation of Cao and Wei's methodology [**30**] and additionally Davis' methodology[**31**] for generating synthetic Berlin daily maximum temperature beyond the scope of physical weather forecasts. Finally, a 'weather like' process that is used as a test data set for the new weather generator (presented in chapter 5), is given in section 3.

1. Description of the real temperature data used in this thesis: Summary statistics

This part of the chapter describes the real weather data that is later used in various experiments.

1.1. Germany, Berlin: daily maximum temperature. Consider daily maximum temperature data for Berlin, Germany. The time range covered by the data is: 1876/01/01 - 2005/08/01, and there are only 2 isolated and 2 consecutive values missing in the time series. A section of the data set is illustrated on figure 1.

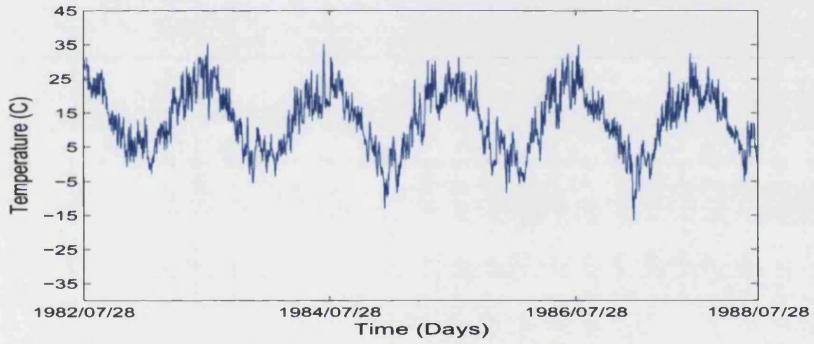


FIGURE 1. Daily maximum temperature for Germany, Berlin, for the period: 1982/07/28 - 1988/01/08

Table 1.1 displays the summary statistics of the Berlin daily maximum temperature.

For the Berlin maximum daily temperature the relative frequency is presented in figure 2, together with 95 % and 5% percentiles (red).

We also present the distribution of extreme values in figure 3, where extreme values are defined to be values observed above the 95 % and below 5 % percentiles of the total data set.

Mean	13.053
Standard Deviation	9.0177
Median	13.1
Kurtosis	2.2052
Skewness	-0.0379
95 % Percentile	27.3
5 % Percentile	-0.9
Average number of freezing days per year[1]	23.628

TABLE 1. Summary statistics of daily Berlin maximum temperature based on 1982/07/28 - 1988/01/08 historical data.

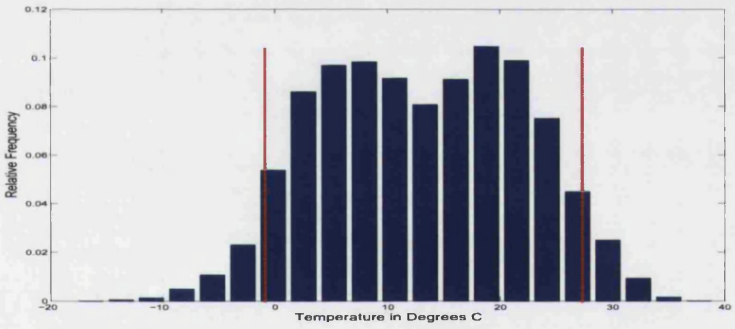


FIGURE 2. Relative frequency distribution of all observed maximum daily temperatures for Germany, Berlin, for the period: 1876/01/01 - 2005/08/01 together with 95 % and 5 % percentiles presented in red.

The relative histograms of first and second differences is presented in figure 4, together with the 95 % and 5 % percentiles (red). Additionally table 1.1 contains various percentiles values for the first and second differences.

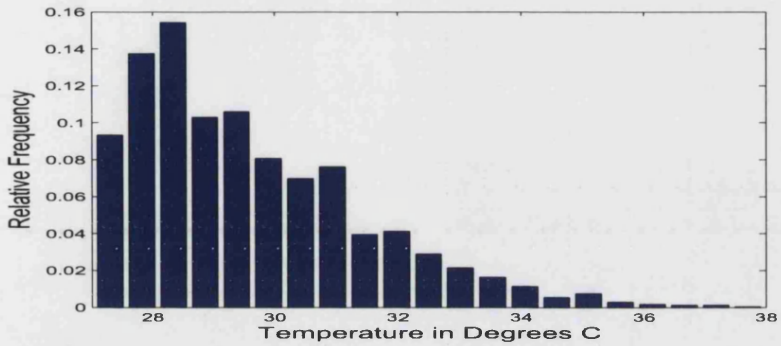


FIGURE 3. Relative frequency of observed extreme high observations (above 95 % percentile of Berlin daily maximum data.

First and second differences D_t^1 D_t^2 are defined as presented in equations 3 and 4 respectively.

$$D_t^1 = s_t - s_{t-1}, \tag{3}$$

$$D_t^2 = s_t - s_{t-2}, \tag{4}$$

where s_t is temperature observed at time t.

The relative frequency of extreme first differences is given in figure 5 (again extreme is defined to be values in the tail of observed relative frequency, i.e. below 5 % and above 95 %).

In section 3 we construct a data set, the components of which behave like certain characteristics of temperature data. In particular, the main characteristics include: cold and warm fronts, seasonality and daily oscillations. A *front* is defined to be a temperature pattern where a sudden (over 1 or 2 days) 'significant' rise (for warm) or drop (for cold) in temperature occurs.

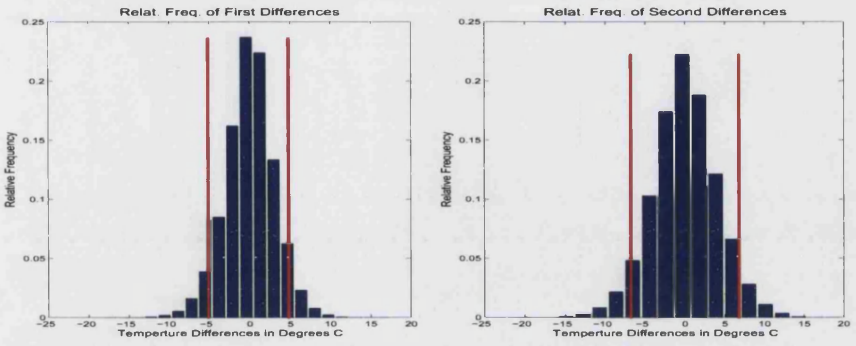


FIGURE 4. Relative frequency of first and second differences, observed in Berlin data, together with the 5 % and 95 % percentiles (red).

Percentile and difference type	**Value in degrees C**
95 % Percentile of First Differences	4.8
5 % Percentile of First Differences	-5.1
97.5 % Percentile of First Differences	6.8
2.5 % Percentile of First Differences	-6.8
95 % Percentile of Second Differences	5.9
5 % Percentile of Second Differences	-6.3
97.5 % Percentile of Second Differences	8.2
2.5 % Percentile of Second Differences	-8.4

TABLE 2. Table of percentiles of first and second differences of the Berlin data.

'Significant' can be defined as a drop/rise in temperature on the level observed in the tails of the relative frequency distribution for first and second differences.

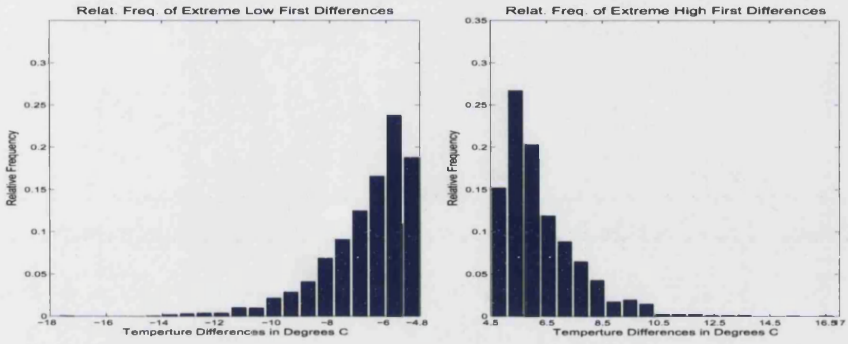


FIGURE 5. Relative frequency of extreme first differences, Berlin data.

Here we present fronts observed in the Berlin daily maximum temperature data at various times of the year. In particular, figure 6 illustrates examples of warm fronts observed in summer; figure 7 displays cold fronts examples observed in winter; and finally figure 8 contains examples of both cold and warm fronts observed in the autumn and spring periods.

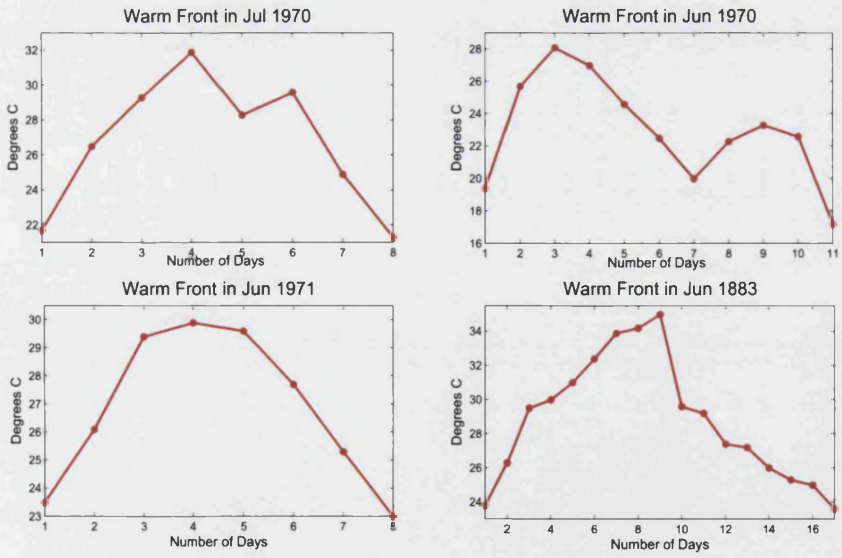


FIGURE 6. Example of warm fronts observed in the summer periods of the Berlin data.

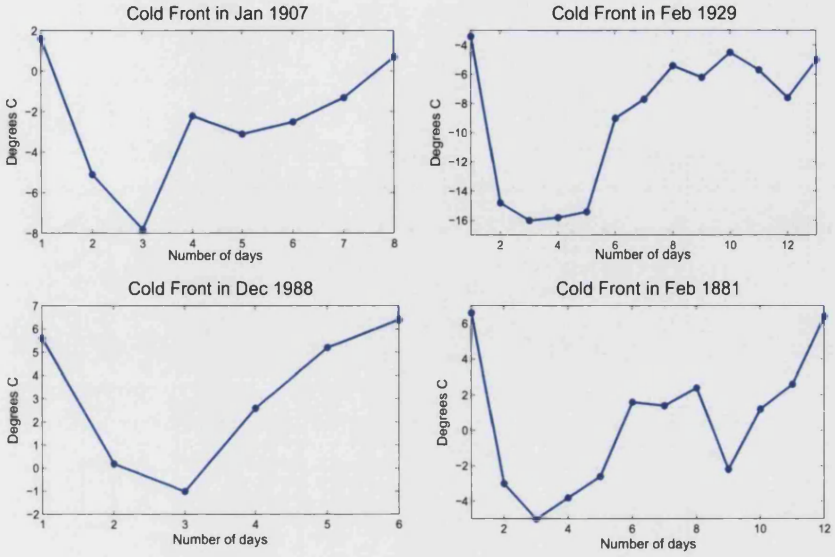


FIGURE 7. Example of cold warm fronts observed in the winter periods in Berlin data.

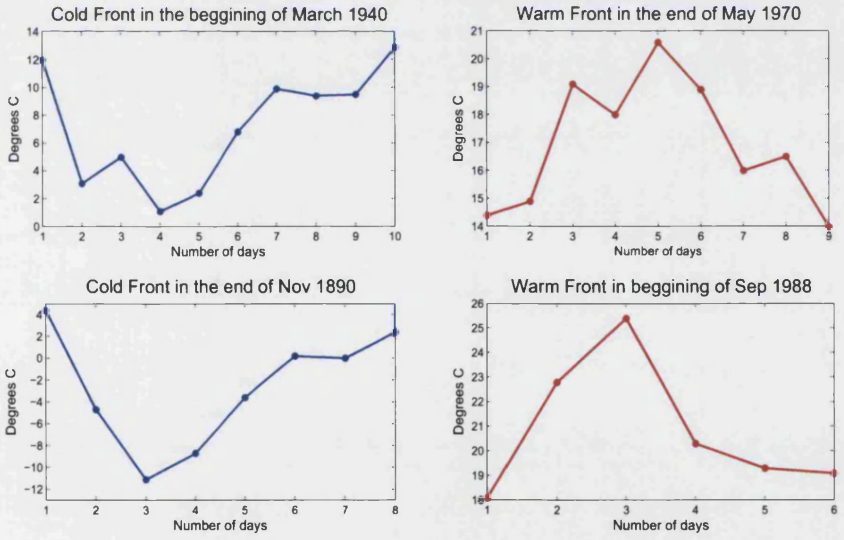


FIGURE 8. Example of cold and warm fronts observed in the spring and autumn periods in Berlin data.

Also we present the relative frequencies of the first and second differences as defined in equations 3(figure 9) and 4 (figure 10) respectively for each

Season	Mean	Standard Deviation
Winter	1.9548	1.7144
Summer	2.6269	2.0866
Spring	2.6146	2.073
Autumn	2.0046	1.6708

TABLE 3. Mean and variance statistics for the absolute first differences for different seasons.

season separately, presented together with the sample mean and variance statistics. This information is later used in construction of the test data set, section 3.

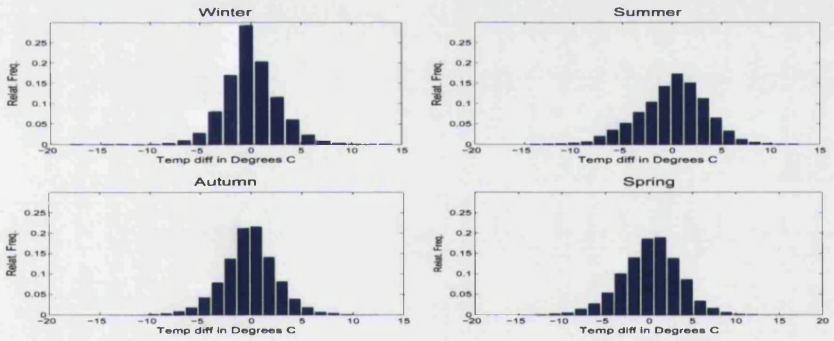


FIGURE 9. Relative Frequency distribution of first differences by season.

Finally, as we model daily oscillations and fronts separately in the 'weather like' process we want to identify the mean and variance of the first and second differences of our data after the fronts have been removed.

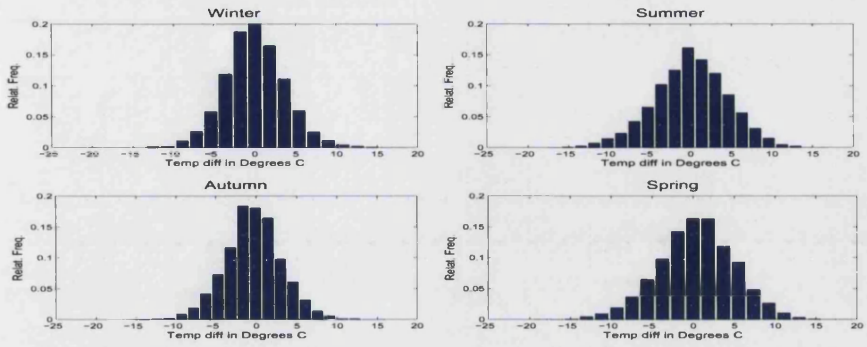


FIGURE 10. Relative Frequency distribution of second differences by season.

Season	Mean	Standard Deviation
Winter	3.8335	2.3378
Summer	3.5287	2.7381
Spring	3.6424	2.832
Autumn	2.7387	2.1944

TABLE 4. Mean and variance statistics for the absolute second differences for different seasons.

2. Generating synthetic weather data: common statistical approaches.

In this section we reproduce synthetic weather data using the methodologies developed by Cao and Wei [**30**] and Davis [**31**]. The generated synthetic data is then used in chapter 6 to price a weather derivative. These prices and various statistics are then compared to prices and statistics that are obtained using historical data alone and the proposed new weather generator.

The temperature modelling approach used by Cao and Wei and Davis are very popular techniques used extensively in the financial industry to

Season and difference	Mean	Standard Deviation
Winter First Difference	1.8937	1.6093
Summer First Difference	2.5922	2.0358
Summer Second Difference	3.3335	2.5399
Winter Second Difference	2.6092	2.0791

TABLE 5. Mean and standard deviation of the first and second difference after the removal of fronts as identified in this section.

model weather, particular temperature. Hence it is important to use those techniques as a benchmark for the new methodology (chapters 5 and 6).

Cao and Wei [**30**] model temperature by trying to capture several characteristics of real temperature data such as: 'seasonal cyclical patterns'; 'daily variations in temperature around some 'normal' temperature'; 'autoregression property of temperature changes'; 'seasonal extent of variations (bigger in the winter and smaller in the summer)', [**30**]. Davis, [**31**] proposes similar, but simpler structure to model temperature time series, using also 'normal' temperature for daily variation and an autoregressive [**26**] process as a simple temperature change approximation. The methodology of both is implemented using the Berlin data (data description is given in section 1). Now lets consider each methodology, in a little more detail.

We used Berlin data from 01-Jan-1983 to 31-Dec-2002 to produce synthetic temperature time series according to Davis' methodology [2]. Let the daily temperature series be denoted $\{T_i, i = 1, \ldots 4015\}$, and $\{\overline{T_j}, j = 1, \ldots 365\}$

[2] 20 years were used as calibration period by Cao and Wei

denotes the long term average temperature for each day of the year. $\overline{T_j}$ is obtained by computing an average between 20 temperatures on a corresponding date (for example, average 1st of Jan, average 2nd of Jan and so on) and then smoothing the series using a moving average, with 9 points used in the construction of the smoothed average. Figure 2 illustrates an example of one year of the Berlin data (1984) and the resulting $\overline{T_j}$.

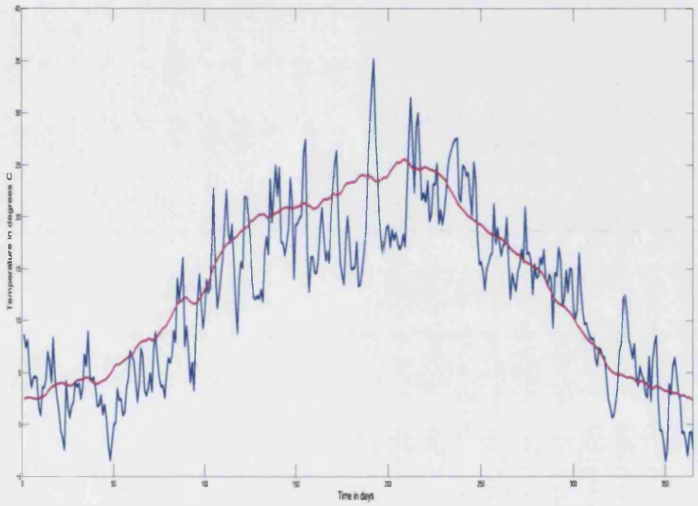


FIGURE 11. SNT (in pink) of Berlin data and the actual observations (blue) for the 1984.

Then the deviation, denoted as $D_i = T_i - \overline{T_j}$, (where j is the corresponding day of the year to i) is modelled as a Autoregressive Process of order one [**26**]:

$$D_i = a_1 D_{i-k} + c_1 + b\epsilon_i, \tag{5}$$

where ϵ_i are independently identically distributed normal random variables [**25**] with mean zero and variance one and a_1, b_1 and c_1 are constants, which are determined using a least squares [**25**] optimisation method. For the Berlin data the optimal parameter values, according to least squares were: $a_1 = 0.01$, $b_1 = 2.8685$ and $c_1 = 0.0015$. Examples of 250 generated temperature paths are given in figure 2 together with the actual 2003 Berlin temperature observations.

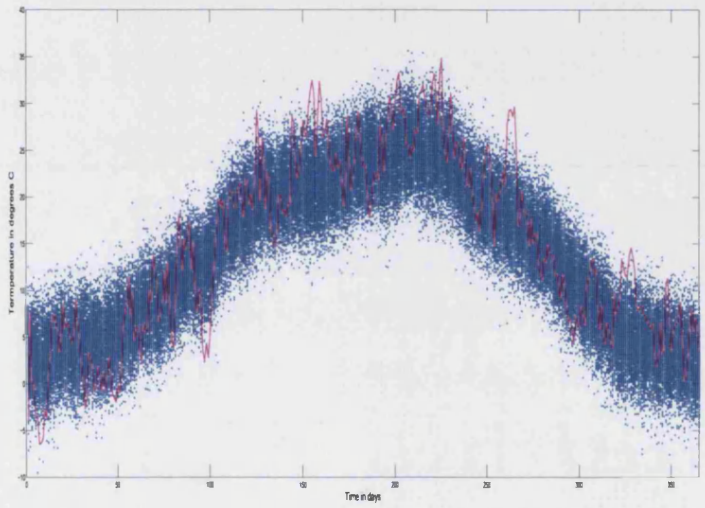


FIGURE 12. Generated 250 paths of synthetic weather data according to Davis methodology (blue) and the actual 2003 Berlin data (pink).

For the implementation of Cao and Wei model we again used Berlin data from 01-Jan-1983 to 31-Dec-2002. Cao and Wei's model is defined as following. Let yr index the years in the sample period, such that $yr = 1$ for 1984, $yr = 2$ for 1985, etc. Also let d index days in a year, such that $d = 1$

is 1st of Jan, $d = 2$ is the 2nd of Jan and so on. Denote temperature on day d in year yr as $Y_{yr,d}$. Then the generated temperature $Y_{yr,d}$ is given by:

$$Y_{yr,t} = \widehat{Y}_{yr,t} + U_{yr,t}, \tag{6}$$

where $U_{yr,t}$ is the daily temperature residual, which follows a $k = 3$ lag autoregressive process. In particular,

$$\begin{aligned} U_{yr,t} &= \sum_{i=1}^{k} (\rho_i U_{yr,t-i}) + \sigma_{yr,t} * \epsilon_{yr,t}, \\ \sigma_{yr,t} &= \sigma_0 - \sigma_1 \left| sin\left(\pi t/365 + \phi\right) \right|, \\ \epsilon_{yr,t} &\sim N(0,1). \end{aligned} \tag{7}$$

where ρ_i,σ_0,σ_1,ϕ are parameters of the model.

$\widehat{Y}_{yr,t}$ is the mean and the trend of the time series. It is computed, by first taking an average over all the years for a particular day of a year. For example, average 1st of Jan (average over 20 years of the historical data), average 2nd of Jan and so on. These daily averages are often referred to as Seasonal Normal Temperature (SNT). Then monthly averages based on SNT can be determined, and are referred to as monthly SNT.

Cao and Wei further propose to adjust 20 year daily SNT in accordance with the deviation of a particular month in a particular year from the monthly SNT. In particular, they compute:

(1) For each month the average of 20 year based daily averages (Traditional SNT) -12 such averages;

(2) For each particular year they compute the realized average temperature of each month;

(3) For each month, they find the difference between the actual monthly average from Step 2 and the the average from Step 1;

(4) Finally, for each day of the month, they adjust the historical SNT by the quantity calculated in Step 3.

This series is denoted as Adjusted SNT. Cao and Wei's approach to use Adjusted SNT to construct $\widehat{Y}_{yr,t}$ produces a better model fit to the historical data in terms of Maximum Likelihood [**25**]. It only works, however, for the historical data. The same technique can not be used to extrapolate forward, as the monthly deviations from 20 year based SNT are not known in advance. Hence, it makes more sense to use SNT values to represent $\widehat{Y}_{yr,t}$. Here we generate temperature time series using both SNT and Adjusted SNT. And in chapter 6 we compare prices generated using both SNT and Adjusted SNT. Figure 2 illustrates actual Berlin temperature (blue) for 1984 the generated SNT (in pink) and the Adjusted SNT (in red).

By construction, a simulated temperature $Y_{yr,t}$ at time t, yr (see equations 6 and 7) is a Normally distributed random variable with mean

$$E\left(Y_{yr,t}\right) = \widehat{Y}_{yr,t} + \sum_{i=1}^{3} \left(\rho_i U_{yr,t-i}\right), \tag{8}$$

and variance

$$\sigma^2_{yr,t} = \left(\sigma_0 - \sigma_1 \left|sin\left(\pi/365 + \phi\right)\right|\right)^2. \tag{9}$$

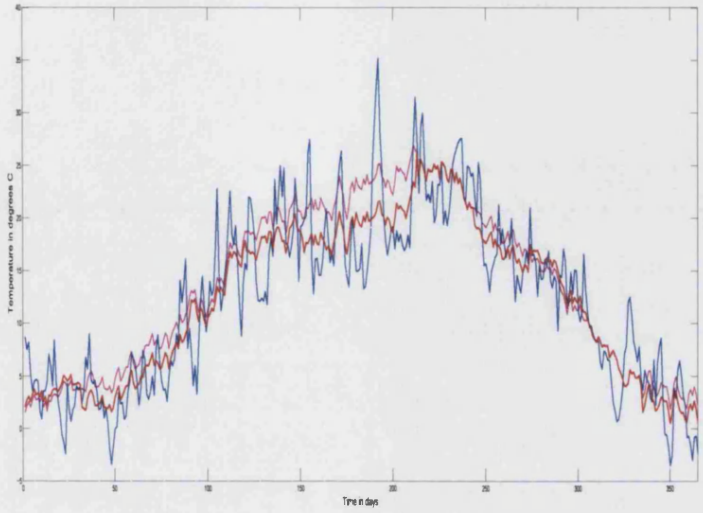


FIGURE 13. SNT (in pink), Adjusted SNT (in red) produced according to Cao and Wei methodology of Berlin data and the actual observations (blue) for the 1984.

The set of parameters $\Theta = (\rho_1, \rho_2, \rho_3, \sigma_0, \sigma_1, \phi)$ is assumed to be optimal when the likelihood function is maximized. In order to simplify the problem the log-likelihood is used instead. The log-likelihood function is given in equation 10

$$l(\Theta; Y) = -\frac{1}{2} \sum_{yr=1}^{20} \sum_{t=1}^{T} \left(\frac{[Y_{yr,t} - E_{yr,t-1}(Y_{yr,t})]^2}{\sigma_{yr,t}^2} + ln\left(2\pi\sigma_{yr,t}^2\right) \right) \tag{10}$$

The parameters are optimal when log-likelihood function is maximized subject to:

$$(\sigma_0 - \sigma_1 |sin(\pi/365 + \phi)|) > 0 \tag{11}$$

Season	SNT based	Adjusted SNT based
ρ_1	0.80321253	0.823964237
ρ_2	-0.093194699	-0.083310864
ρ_3	0.074565646	0.069397676
σ_0	3.489280269	3.74983828
σ_1	1.121357784	1.177412501
ϕ	1.678540421	1.583706887

TABLE 6. Table of parameters of the Cao and Wei model for the Berlin data.

The table of parameters for the Berlin data optimized using maximum Likelihood is presented in table 2

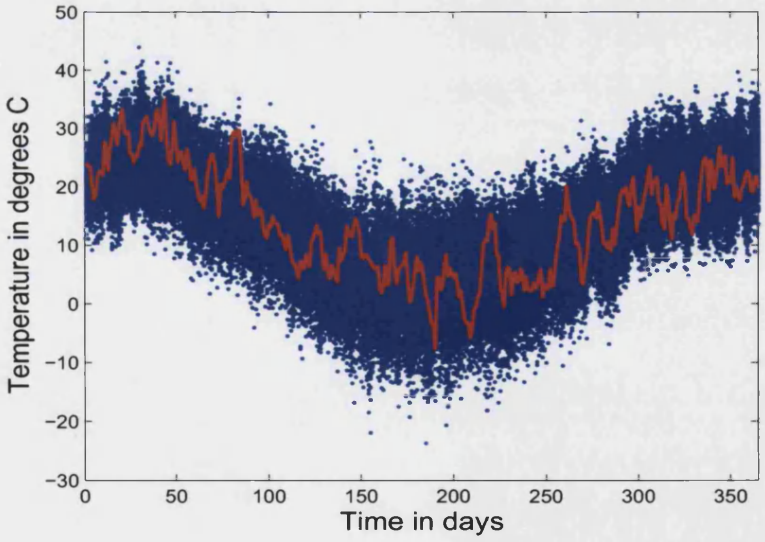


FIGURE 14. Generated 250 paths of synthetic weather data according to Cao and Wei methodology (blue) and the actual 2003/2004 Berlin data (red).

3. Test data: Synthetic weather data - characteristics and parameters

One of the main aims of this thesis is to compare synthetic weather data, produced by popular weather generators to synthetic weather data produced by the proposed new weather generator (chapter 5), when used for financial risk management and weather derivative pricing. In general, whether the interest lies in hedging weather derivatives other areas of risk management, or perhaps different application areas such as logistics, health, tourism, transport etc, estimates of future weather over a long time period are required. Here we propose a process that mimics the behaviour of temperature data, which is then used to test the proposed new weather generator in a controlled experimental environment.

3.1. Overall structure of the mathematical model. The discrete weather-like process is of the form:

$$\begin{aligned} X(t) &= F\left(A\left(t\right),\phi\left(t\right),t\right)+\Upsilon\left(\Phi\left(t\right),t\right)+ \\ &+ \Lambda\left(\Phi\left(t\right),t-t_c^*\right)\Gamma(\Phi\left(t\right),t)\Im_c\left(t-t_c^*\right) \\ &\quad +\Lambda\left(\Phi\left(t\right),t-t_w^*\right)\Gamma(\Phi\left(t\right),t)\Im_w\left(t-t_w^*\right)+\kappa+\xi \end{aligned} \tag{12}$$

This weather-like process consists of six terms: a periodic term $F\left(A\left(t\right),\phi\left(t\right),t\right)$, which represents seasonality of the real weather data; a random autoregressive term $\Upsilon\left(\Phi\left(t\right),t\right)$, which represents daily fluctuations of the real weather data; a term $\Lambda\left(\Phi\left(t\right),t-t_c^*\right)\Gamma(\Phi\left(t\right),t)\Im_c\left(t-t_c^*\right)$ which represents

cold fronts, that are observed in winter in the real weather data (see chapter 1); a term $\Lambda(\Phi(t), t - t_w^*)\,\Gamma(\Phi(t), t)\Im_w(t - t_w^*)$ which represents warm fronts that are observed in summers in the real weather data (see section 1); a constant term κ that controls the mean of the output data; and finally a white noise term ξ_t which represents measurement and operational error. t is time (in days), $\phi(t)$ represents small shifts which affect the periodicity of the seasonal component, $\Phi(t)$ is the annual phase of the generated year, where $\Phi(t) = \frac{1}{365}, \frac{2}{365}, \ldots, \frac{365}{365}$, and finally $t - t_c^*$ and $t - t_w^*$ are times since onset of fronts. The coefficients, summarised in the table 7 are discussed below; they can be tuned such that the output data has the characteristics similar to the climate of a preferred chosen location.

Each term is now considered separately in more detail. First consider the periodic term $F(A(t), \phi(t), t)$, which is given by equation 13. The tuning of the parameters of this part of the process is driven by the seasonality observed in the daily temperature data.

$$F(A(t), \phi(t), t) = A(t)\, sin\,(w_1 t + \phi(t)) \tag{13}$$

The seasonality structure of temperature data is chosen to be represented in functional form as a sine wave. Where $A(t)$ is an amplitude, w_1 defines the period of the sine wave (such that the period of sine wave is equal to 365 simulated days), and $\phi(t)$ controls small shifts in the length of the period.

Now examine each part of the seasonal component in even more detail. The amplitude $A(t)$ is given by equation 14.

(14) $$A(t) = A_0 (1 + \varrho sin(w_3 t))$$

where ω_3 is a constant that defines the period of the variation in the amplitude and ϱ is a scaling constant that controls the impact of the time varying component of the amplitude of the seasonal factor of the generated 'weather like' data.

The amplitude is a function of time, constructed such that it exhibits small fluctuations around A_0 as time increases. A_0, which is also given in table 7, is chosen such that the time series produced will have an average temperature spread consistent with that observed in the chosen climate. The amplitude of the sine wave is not the only parameter that was chosen to vary with time. The periodicity of the sine wave does not remain the same for each generated year. This allows us to produce a realistic looking temperature series which can be compared to real temperature data. It also allows us to keep a sufficient level of complexity in the generated process. That is achieved by setting $\phi(t)$ as it is given in equation 15.

(15) $$\phi(t) = \eta \cos(\omega_2 t)$$

where ω_2 specifies the period and η is a scaling constant that controls the impact of the time changing component on the periodicity shift of the seasonal component.

Both the amplitude $A(t)$ and the periodicity shift $\phi(t)$ are illustrated in figure 15. The complete seasonal component $F(A(t), \phi(t), t)$ of the generated data is shown in figure 16.

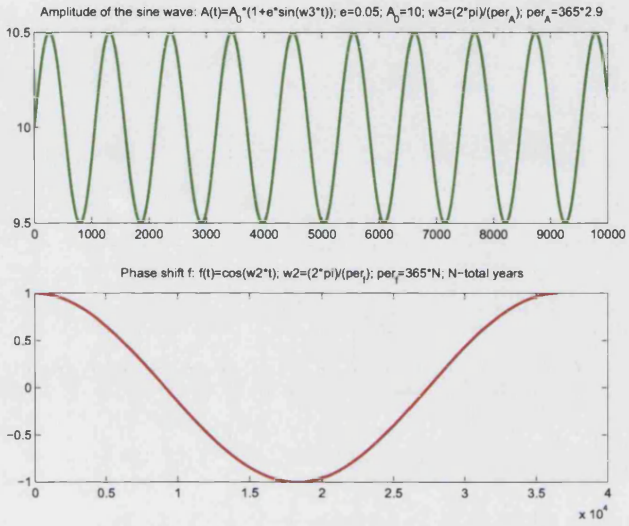


FIGURE 15. The amplitude (top plot) and the shift (bottom) of the seasonal component of the process.

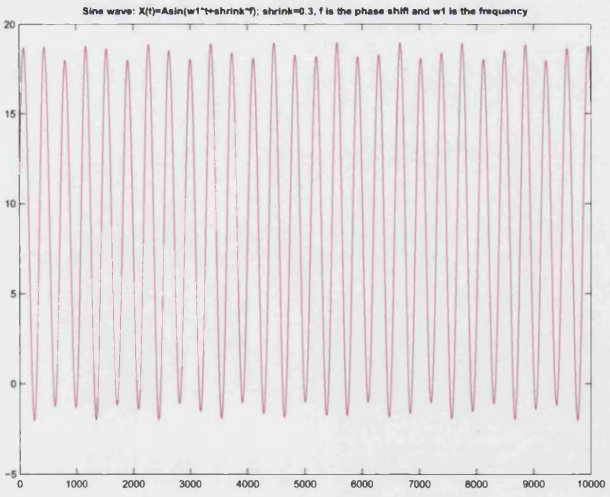


FIGURE 16. Periodic signal $F(A(t), \phi(t), t)$ of the generated process.

Note that the chosen value of ω_3 creates an approximately three year periodic shift in the amplitude of the seasonal component. The chosen ϱ results in an impact on the time varying factor in the amplitude of the seasonal component ranging from -0.5 to 0.5. Finally, the value of ω_2 impacts the periodicity shift of the seasonal component in a smooth decreasing and then increasing fashion over a period of 40,000 days. This is an artificially chosen behaviour that is not necessarily present in the real data. Identifying the periodicity of the actual weather data is a study in itself, and is not investigated in this thesis. It is also not clear whether the changing periodicity of the actual weather data can be defined in a functional form. These components were chosen for the purpose of creating a complex enough test data set, with varying length periodic patterns, and behaves like weather data, rather than creating 'the best model' to model actual weather time series. In particular, we wanted to create a data set that contains patterns on various time scales to see how the new proposed weather generator (chapter 5) performs on a non-linear, weather like process, that is seasonal and periodic, but periodic on multiple time scales.

The second term of the process, $\Upsilon(\Phi(t), t)$, represents random daily fluctuations (or in other words volatility [**24**]) observed in temperature data. $\Upsilon(\Phi(t), t)$ is an Autoregressive order 3 (AR(3), [**26**]) process which is given by equation 16.

$$\Upsilon(\Phi(t), t) = \Psi(\Phi(t), t)(a\Upsilon_{t-1} + b\Upsilon_{t-2} + c\Upsilon_{t-3} + \varepsilon_t) \tag{16}$$

where ε_t is iid[3] and $\varepsilon_t \backsim N(0,1)$ and represents noise of the AR(3) process and a, b, c are constants of the AR(3) process.

The amplitude of the AR(3) process, $\Psi(\Phi(t), t)$, is a function of time t but is also a function of the phase of the year $\Phi(t)$. In the actual Berlin data set (see section 1) it has been observed that the volatility (in terms of first differences) varied through the year. In order to mimic this behaviour pattern in the weather-like process, the random oscillations that are produced by the AR(3) process are chosen to have different amplitudes for the different seasons of a year, equation 17.

$$\Psi(\Phi(t), t) = \begin{cases} \tilde{a} + \delta_1 \cos^2(\omega_4 t), & \text{otherwise} \\ \tilde{a} + \delta_2 \cos^2(\omega_4 t), & \Phi(t) \in \left[\frac{30}{365}, \frac{122}{365}\right] \\ \tilde{a} + \delta_3 \cos^2(\omega_4 t), & \Phi(t) \in \left[\frac{210}{365}, \frac{290}{365}\right] \end{cases} \tag{17}$$

where $\tilde{a}$ is a constant that affects the mean of the amplitude. δ_1, δ_2 and δ_3 are the magnitudes of the impact of the time changing component. A change to these parameters would create larger or smaller volatility in the weather-like process. ω_4 is the periodicity of the amplitude of the AR(3) process.

Again, ω_4 and δ_1, δ_2, δ_3 were chosen arbitrarily to add complexity to the generated test 'weather like' process. The choice of ω_4 results in the amplitude of the AR(3) process being periodic, with a period of 1800 days. Cao and Wei [**30**] (section 2) assumed a seasonal structure for the volatility, where winter was more volatile compared to summer. We have seen that, in

[3]Independent Identically Distributed [**25**]

terms of mean absolute first difference, this was not the case for the Berlin data (see section 1). Although the severity of volatility observed in different seasons depends on the measure of volatility that is used (such as max, max of absolute values, mean, mean of absolute values, s.d. of actual or absolute values, of first/ second/third differences), in order to pick δ_1, δ_2 and δ_3 we will concentrate on second differences. According to the mean of the second differences, winter is the most volatile, followed by spring, summer and finally autumn, see table 1.1.For simplicity spring and autumn have chosen to have the same volatility. The second term $\Upsilon(\Phi(t), t)$ of the weather-like process and its amplitude are shown on the figure 17.

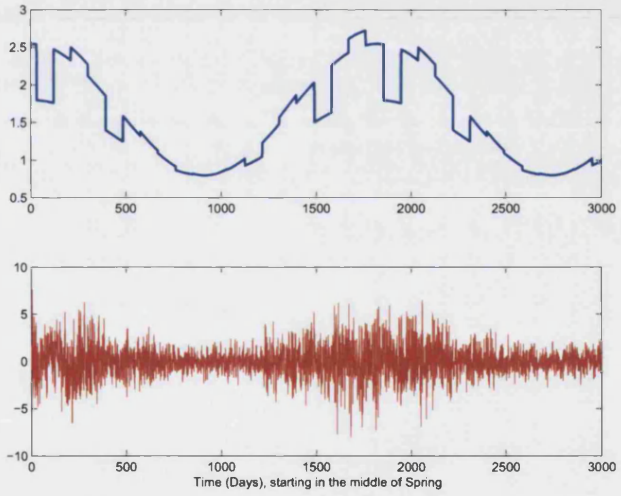


FIGURE 17. $\Upsilon(\Phi(t), t)$ (bottom plot) and its amplitude $\Psi(\Phi(t), t)$ (top plot).

The third term of the process represents the effects of cold fronts. In a typical cold front the temperature drops suddenly by several degrees and then takes several days to recover (see section 1, figures 7 and 8). The template for the synthesised cold fronts is illustrated in the top plot of

figure 18; it is given by the equation 18. $\Im_c(t - t_c^*)$ has been constructed as a function of time since initialisation of the cold front, $t - t_c^*$.

(18)

$$\Im_c(t - t_c^*) = \begin{cases} \left(n_1(t - t_c^*)^2 - n_2(t - t_c^*)\right) \exp\left(n_3(t - t_c^*)\right), \\ \text{for } t - t_c^* = 1, 2, \ldots 12; \\ 0, \qquad \text{otherwise.} \end{cases}$$

where n_1, n_2, and n_3 are constants that affect shape of the function and $t - t_c^*$ is the time since onset (initialisation of a cold front) with $t - t_c^* = 1 \ldots 12$ days.

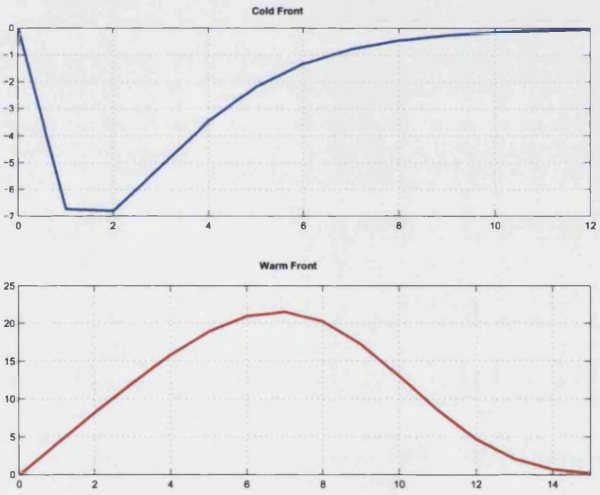


FIGURE 18. Generated cold (blue) - top plot and warm (red) - bottom plot fronts. x-axis represents time in days and y axis represents temperature in degrees C.

Similarly, the fourth part of the process represents warm fronts in the weather data. The main feature of a warm front is that the temperature rises by several degrees over a couple of days and, after reaching its peak

temperature, drops slowly - sometimes staying at the peak for several days (section 1, figures 6 and 8). A synthesised warm front, shown on the bottom of figure 18, is given by equation 19. As in the case of cold fronts $\Im_w(t-t_w^*)$ is constructed as a functions of time since initialisation, $t-t_w^*$ (measured in days),

(19)

$$\Im_w(t-t_c^*) = \begin{cases} \left(m_1(t-t_w^*)^2 + m_2(t-t_w^*)\right)\exp\left(m_3(t-t_w^*)^4\right), \\ \text{for } t-t_w^* = 1,2,\ldots 15; \\ 0, \qquad \text{otherwise.} \end{cases}$$

where m_1,m_2,m_3 are constants.

Both $\Im_c(t-t_c^*)$ and $\Im_w(t-t_w^*)$ are episodic in their nature and are chosen not to occur in some parts of the year. During the rest of the year, in particular from June to August plus or minus 9 days of autumn and spring (for warm fronts), and from December to February plus minus 15 days of autumn and spring (for cold fronts), the probability of both warm and cold front occurring are for simplicity, chosen to be equal and are denoted as p_f. In order to define the initialisation of cold or warm fronts a function $\Lambda(\Phi(t), t-t^*)$ is created using equation 20. $\Lambda(\Phi(t), t-t^*)$ is a function of the phase of the year $\Phi(t)$ and time since onset.

(20)

$$\Lambda(\Phi(t), t-t^*, t) = \begin{cases} 1, & rand \leqslant p \text{ and } \left(\Phi(t) \in \left[\frac{20}{365}, \frac{130}{365}\right] \text{ or } \Phi(t) \in \left[\frac{200}{365}\ \frac{310}{365}\right]\right) \text{ and } t-t^* = 0 \\ 0, & \text{otherwise.} \end{cases}$$

and *rand* is a random number drawn from a uniform distribution between zero and one.

When $\Phi(t) \in \left[\frac{20}{365} \ldots \frac{130}{365}\right]$, $t - t_w^* = 0$ and $rand \leq p_f$ a warm front is initialised when $\Phi(t) \in \left[\frac{200}{365} \ldots \frac{310}{365}\right]$, $t - t_c^* = 0$, $rand \leq p_f$ a cold front is initialised. Once a front has been initialised it goes through its full cycle (i.e. 12 time steps for a cold front and 15 time steps for a warm front).

The functional form of the fronts in the weather-like process, and their occurrences, have been tuned to match the behaviour of real fronts in the actual weather data. In order to do so we use the definition of fronts from section 1, additionally specifying that cold fronts are characterised by a temperature drop of more than approximately 6 degrees over one or two days, which corresponds to 2.5th % percentile of the first differences (table 1.1) in the Berlin data. Similarly, we define warm fronts to be characterised by an approximately 8 degree rise over one or two days, which corresponds to 97th % percentile of second differences in the Berlin data(table 1.1). This construction allowed extra flexibility by considering drop (cold fronts) and rises (warm fronts) using first and second differences, in order to account for the fact that in the actual Berlin data sharp rises/falls sometimes occur over a longer period of time than one day (see figures 6, 7 and 8). Then, the chosen functional form for fronts is a smoothed, simplified representation of the behaviour actually observed in the data, some examples of which are represented in section 1

In order to propose a reasonable guess for the probability of occurrence and the amplitude of fronts, we study the occurrence and severity of fronts

in the Berlin data. First, we compute fronts that happened based on the first difference (182.6 cold fronts and 77 warm fronts occurrences) and the number of fronts that happened based on the second difference (700 cold fronts and 429 warm fronts occurrences). From this we compute the yearly average occurrence of fronts according to the first (1.46 cold fronts and 0.598 warm fronts per year on average) and second differences (5.42 cold fronts and 3.3316 warm fronts per year on average), by dividing by the total number of years in the available Berlin record. We then compute the average between the first and the second difference based average yearly occurrence of fronts (3.418 cold fronts and 1.96 of warm fronts). Finally, based on the observed average yearly occurrence of fronts in the chosen period one can approximate the occurrence of fronts to be:

$$p_f = \frac{\text{Yearly\_ave\_occur\_fronts}}{\text{Max\_fronts\_per\_year}}, \tag{21}$$

where

$$\text{Max\_fronts\_per\_year} = \frac{\text{Total\_nu\_of\_days\_fronts\_may\_occur}}{\text{length\_ of\_front}} \tag{22}$$

Yearly_ave_occur_fronts is yearly average occurrence of fronts;

Max_fronts_per_year is the maximum possible number of fronts per year; Total_nu_of_days_fronts_may_occur is the total number of days where fronts

may occur; and length_ of_front is a length of a front as defined by its functional form. Hence, for cold and warm fronts sampled probability of occurrence is equal to:

$$p_f = \begin{cases} \frac{3.419}{10} = 0.34, & \text{for cold fronts} \\ \frac{1.965}{7.35} = 0.267, & \text{for warm fronts} \end{cases} \tag{23}$$

As we have chosen to use one probability of occurrence for both warm and cold fronts, approximately an average between those probabilities suffices.

In the real temperature data the amount by which temperature drops (cold front) and rises (warm front) would vary depending on the time of the year. In order to account for these behaviour patterns the function $\Gamma(\Phi(t), t)$, which is given by the equation 24, is introduced into the weather-like process. It is the amplitude of fronts and it changes for different time periods in the generated winter and summer months. $\Gamma(\Phi(t), t)$ is illustrated in figure 19 and is a function of time and the phase of the year. The parameters of the chosen functional form for this amplitude were roughly based on the amplitudes of fronts observed in the the Berlin data. Again, it is important to note that the 'weather like' process is constructed as a test data set, rather than a realistic model. For that reason, the 20th % percentile of the actual front amplitudes, observed in the Berlin data was chosen to represent the maximum amplitude of generated fronts. This allows us to generate 'dramatic' drops/rises that are not too unrealistic compared to the actual data.

(24)

$$\Gamma(\Phi(t),t)=\begin{cases} 0, & \Phi(t)\in\left[\frac{0}{365},\frac{20}{365}\right] \text{ or } \Phi(t)\in\left[\frac{130}{365}\ \frac{200}{365}\right] \text{ or } \Phi(t)\in\left[\frac{310}{365}\ \frac{365}{365}\right] \\ s_1, & \Phi(t)\in\left[\frac{20}{365},\frac{30}{365}\right] \text{ or } \Phi(t)\in\left[\frac{120}{365}\ \frac{130}{365}\right] \\ s_2, & \Phi(t)\in\left[\frac{200}{365},\frac{210}{365}\right] \text{ or } \Phi(t)\in\left[\frac{300}{365}\ \frac{310}{365}\right] \\ s_3, & \Phi(t)\in\left[\frac{30}{365},\frac{60}{365}\right] \text{ or } \Phi(t)\in\left[\frac{90}{365}\ \frac{120}{365}\right] \\ s_4, & \Phi(t)\in\left[\frac{60}{365},\frac{90}{365}\right] \\ s_5, & \Phi(t)\in\left[\frac{210}{365},\frac{240}{365}\right] \text{ or } \Phi(t)\in\left[\frac{270}{365}\ \frac{300}{365}\right] \\ s_6, & \Phi(t)\in\left[\frac{240}{365},\frac{270}{365}\right] \end{cases}$$

The fifth term of the weather-like process is a constant, which affects the mean of the output data, the value of which is given in the table 7. This parameter allows us to match the daily min, max or average temperatures in a chosen climate in a simplistic way. Figure 22 illustrates for example the generated 'weather like' process that was shifted using this parameter to match the Berlin daily minimum data record.

Finally, the last term of the process ξ_t, given on the figure 20, is an independently identically distributed normal random variable, with $N(0, 1.5)$, which represents the measurement and operational observational noise present in the temperature data.

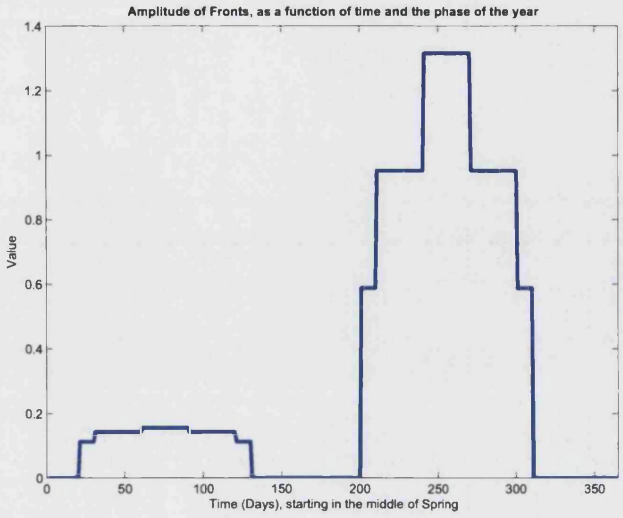


FIGURE 19. Amplitude of fronts, as a function of time and the phase of the year

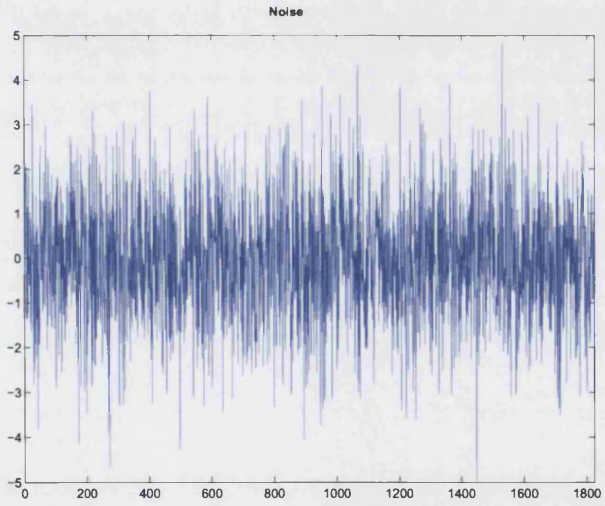


FIGURE 20. Generated random numbers from a Normal distribution with varience=1.5 and mean=0

Our toy weather process can now be used to generate a large amount of data that will form one of the learning sets used to test the proposed new weather generator (chapter 5) in a controlled environment. The generated

data exhibits characteristics of real weather data. The data generated by the process shown in figure 21.

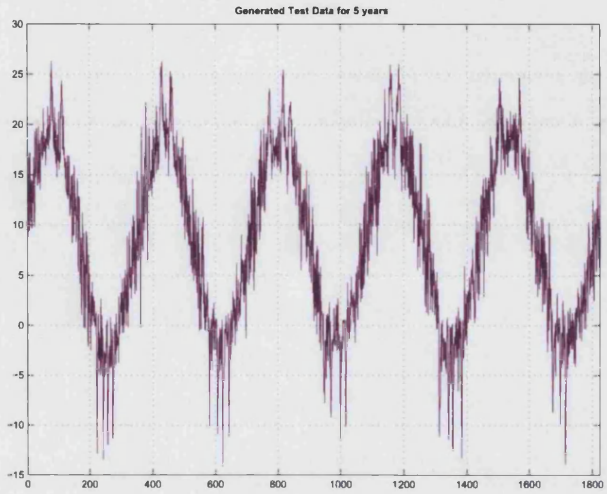


FIGURE 21. Generated toy weather process, zoomed into the five year period.

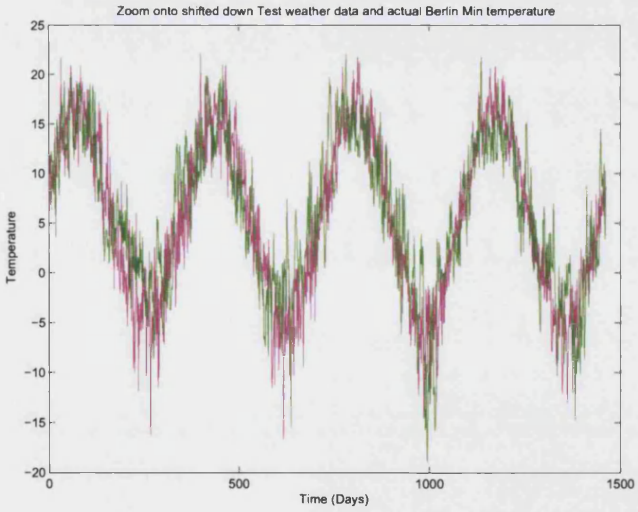


FIGURE 22. Generated toy weather data (pink), together with the min daily temperature, Berlin (green).

All the parameters that have been used in the construction of the weather-like process are summarised in table 7.

Table of Parameters of the Test Weather Data		
N	Number of years generated.	3000
n_1 n_2 n_3	Parameters of the template function that generates a cold front.	$\frac{2}{9}$ $-\frac{40}{3}$ $-\frac{2}{3}$
m_1 m_2 m_3	Parameters of the template function that generates a cold front	$-\frac{3}{144}$ $\frac{50}{12}$ $-\frac{2.3}{12^4}$
$\breve{N}_w$	Number of points (days) that the template function of a warm front generates.	15
$\breve{N}_c$	Number of points (days) that the template function of a cold front generates.	12
$p_c(\phi(t))$	Probability of a cold front occurring on a given day in the winter season.	0.3
$p_w(\phi(t))$	Probability of a warm front occurring on a given day in the winter season	0.3
w	Extra days outside the summer season where fronts may occur.	10
$\widetilde{u}$	Extra days outside the winter season where fronts may occur.	15

Table of Parameters of the Test Weather Data		
s_1		$\frac{0.28*2.5}{6.2}$
s_2	Parameters regulating the amplitude of cold and warm fronts.	$0.28x2.1$
s_3	These parameters change depending on the time of year.	$\frac{0.28x2.5}{4.9}$
s_4		$\frac{0.28x2.5}{4.5}$
s_5		$0.28x3.4$
s_6		$0.28x4.7$
ω_1	Parameter controlling the length of the period of the periodic component.	$\frac{2\pi}{365}$
ω_4	the periodicity of the amplitude of the AR(3) process	$\frac{2\pi}{365}x\frac{1}{10}$
η	The scaling constant that controls the impact of the phase shift on the seasonal term.	0.3
A_0	The amplitude of the amplitude function of the seasonal component (the sine wave).	10
ϱ	Scaling constant that controls the impact of the phase shift in the amp. of the seasonal comp.	0.05

Table of Parameters of the Test Weather Data		
ω_2	Constant that controls the period of the phase of the seasonal component.	$\frac{2\pi}{365N}$
ω_3	Constant controlling the phase of the variation in the amp. of the seasonal component.	$\frac{2\pi}{365*2.9}$
a b c	Constants of the AR(3) process.	0.7 -0.4 0.2
$\tilde{a}$	Constant effecting the mean of the amplitude of the AR(3) process.	0.8
δ_1 δ_2 δ_3	Parameters that control the magnitude of the impact of the time changing component. These parameters create larger or smaller volatility depending on the time of year.	1.75 1 1.95
μ_r	Mean of the Identically Independently Normally distributed random variable.	0
σ_r	Variance of the Identically Independently Normally distributed random variable.	1
k	Parameter that effects the mean of the generated 'weather like' process.	8.5

TABLE 7. Table of parameters of the weather like process.

CHAPTER 4

Ensemble Forecasting: Principles and Practice

In this chapter a method is developed that allows us to produce a long term *forecast*[1] for a given time series, such that it combines information from both: a forecasting model and historical data. The output of a forecasting model is initially in the form of an *ensemble* (see chapter 3). *Kernel-dressing* is used to transform an ensemble forecast into a probability forecast. The main goal of the experiment is to establish and analyse the composition of the forecast, and to determine parameters for the kernel-dressing process.

1. Overview

We are interested in long term forecasting of data produced by what is potentially a *non-linear chaotic system.* The forecast $F^t(y)$ is chosen in a form such that at any given time t it combines information produced by two sources: a *forecasting model* and the *climatology.* The forecasting model is our best guess for the system that generated the data. In other words, a forecasting model attempts to predict the future. The climatology is a set of statistics produced by all data that has been observed so far. The central question is how to combine the information produced by a forecasting model and the information from the climatology for a given *system* and how the chosen combination changes with time. We can also ask a different question

[1]All highlighted words are defined in the Glossary.

that is related to the problem of determining $F^t(y)$: how long does it take for the forecast produced by a forecasting model to lose information for a given *scenario*.

Let's formalise this: assume that a forecasting model at time t produces a forecast of data in a form of probability density function - $p_f^t(y)$; and the climatology produces a historical, frequency-based, forecast - also in the form of a probability density function [**25**], which is denoted by $p_c(y)$. Then at time t the issued forecast $F^t(y)$ will combine the information in those two probability functions and can be represented as shown in equation 25,

$$F^t(y) = \alpha^t p_f^t(y) + (1 - \alpha^t) p_c(y), \tag{25}$$

where α^t is the proportional weight and is set to be: $0 \leq \alpha^t \leq 1$.

We are interested in determining the 'best' forecast $F^t(y)$ with respect to minimising Ignorance [**45**][2]. This poses further questions: how the performance of $F^t(y)$ is evaluated in order to determine the 'best' forecast; how to determine α^t; and how to construct $p_f^t(y)$ and $p_c(y)$. The methodology developed in this chapter was examined in the *perfect model scenario* framework using chaotic systems [**27**] and a threshold autoregressive process [**54**]. The parameters and the methodology of the forecast were established using a *learning set* for each of the studied systems and then applied to produce an *Out of sample* forecast. The definitions of the parameters used in this chapter are presented in table 1.

[2]Definition and explanation of Ignorance skill score is given in section 2.2

N_{ls}	Length of a learning set.
T_e	Lead time covered by an ensemble.
N_{BINS}	Number of bins used in the construction of climatology.
M	Size of an ensemble, i.e. number of members.
ε_t	A random variable [48] representing *Additive noise* on the studied system.
μ_ε	Mean of the distribution [49] of the additive noise.
σ_ε	Standard deviation of the dist.[49] of the additive noise.
s_t	Value of a data point of a learning set at a given location.
η_t	A random variable representing *Perturbation noise.*
μ_η	Mean of the distribution of the perturbation noise.
σ_η	Standard deviation of the dist. of the perturbation noise.
L	Number of randomly chosen locations on a learning set, where data points are perturbed.
T_{MAX}	Maximum forecasting lead time.
μ_i^t	A value of an ensemble member i at each lead time t, where $i = 1 \ldots M$
σ_m^t	A width of a kernel-dressed dist. around a m^{th} ensemble member, i.e. μ_i^t, at a given lead time t.
α^t	The blending parameter between the $p_f^t(x)$ and $p_c(x)$.

TABLE 1. The definitions of parameters: combined forecast experiment.

2. Setting of the experiment.

2.1. Definitions of used standard distributions and numerical schemes. First, let's define some standard probability distributions that are used in this thesis. A Normally distributed [**49**] random variable x with mean μ and standard deviation σ is denoted by $x \sim N(\mu, \sigma)$, where $N(\mu, \sigma)$ is given by:

$$N(\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right). \tag{26}$$

A random variable x is uniformly distributed and is denoted as $x \sim U(a, b)$, where $U(a, b)$ is given by equation 27.

$$U(a,b) = \begin{cases} \frac{1}{b-a}, & for\ a \leq x \leq b; \\ 0, & otherwise. \end{cases} \tag{27}$$

Next let's define the Ignorance skill score [**45**]. In particular, given a *verification set* $y = \{y_1, y_2, \ldots, y_T\}$ the Ignorance skill score is defined by:

$$\langle IGN \rangle = -\frac{1}{T}\sum_{j=1}^{T} \log_2\left(p(y_j)\right), \tag{28}$$

where $p(y)$ is a probability function, y_i is a *verification* point and T is the size of the verification set. Ignorance is a measure that reflects how uncertain you are in your probability forecast $p(y)$ given a verification set [**45**].

Also, let's define the Logit transformation [**26**] of parameters where, for a given parameters $0 < \alpha^t < 1$ and $\sigma_m^t > 0$, one can define new parameters α^{*t} and σ_m^{*t} as presented in equations 29 and 30 respectively:

$$\alpha^{*t} = \ln\left(\frac{\alpha^t}{1-\alpha^t}\right); \tag{29}$$

$$\sigma_m^{*t} = \ln\left(\sigma_m^t\right), \tag{30}$$

such that $-\infty < \alpha^{*t} < \infty$ and $-\infty < \sigma_m^{*t} < \infty$.

Finally, the Runge-Kutta [**47**] numerical integration scheme of fourth order, extended to three dimensions, is defined as follows. For a given system of differential equations with respect to time t:

$$\begin{aligned} \frac{dx}{dt} &= f_x(t,x) \;\; with \; x(t_0) = x_0, \\ \frac{dy}{dt} &= f_y(t,y) \;\; with \; y(t_0) = y_0, \\ \frac{dz}{dt} &= f_z(t,z) \;\; with \; z(t_0) = z_0, \end{aligned} \tag{31}$$

the discrete approximation is given by:

$$\begin{aligned} x_{n+1} &= x_n + \frac{h}{6}\left(k_1^x + 2k_2^x + 2k_3^x + k_4^x\right); \\ y_{n+1} &= y_n + \frac{h}{6}\left(k_1^y + 2k_2^y + 2k_3^y + k_4^y\right); \\ z_{n+1} &= z_n + \frac{h}{6}\left(k_1^z + 2k_2^z + 2k_3^z + k_4^z\right); \end{aligned} \tag{32}$$

where k_i^x, k_i^y, k_i^z with $i = 1, \ldots, 4$ are defined by equation set 33.

$$
\begin{aligned}
k_1^x &= f_x(t_n, x_n),\ k_1^y = f_y(t_n, y_n),\ k_1^z = f_z(t_n, z_n) \\
k_2^x &= f_x\left(t_n + \frac{h}{2}, x_n + \frac{h}{2}k_1^x\right),\ k_2^y = f_y\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_1^y\right), \\
k_2^z &= f_z\left(t_n + \frac{h}{2}, z_n + \frac{h}{2}k_1^z\right); \\
k_3^x &= f_x\left(t_n + \frac{h}{2}, x_n + \frac{h}{2}k_2^x\right),\ k_3^y = f_y\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_2^x\right), \\
k_3^z &= f_z\left(t_n + \frac{h}{2}, z_n + \frac{h}{2}k_2^z\right); \\
k_4^x &= f_x(t_n + h, x_n + hk_3^x),\ k_4^y = f_y(t_n + h, y_n + hk_3^y), \\
k_4^z &= f_z(t_n + h, z_n + hk_3^z).
\end{aligned}
$$

h is a time step

(33)

2.2. Experimental design. The experiment can be divided into several stages:

(1) **Generating a learning set:** First, iterate forward in time a chosen system with a predetermined numerical scheme, such that N_{ls} data points are generated. Add independent realisations of ε, where $\varepsilon \sim N(\mu_\varepsilon, \sigma_\varepsilon)$, to each point of the generated data (other distributions can be considered as possible models for an observational noise). Learning sets are used to *train parameters* of $F^t(y)$. In this thesis, the learning sets will only contain one of the components of the chosen chaotic system - the x component. The methods

used in the production of forecasts, however, could be extended to multi-dimensional data.

(2) **Defining climatology**: The climatology $p_c(y)$ is created by constructing the historical relative frequency, of values in the learning set. The range is split into N_{BINS} equally sized bins, that range from the recorded minimum to the recorded maximum of the learning set.

(3) **Constructing initial conditions:** The predictability at different points in the learning set and, in the case of a chaotic system, across different regions of any attractor, may vary [**50**], and a good forecasting model should account for this. In order to accommodate this phenomena, L randomly selected locations on the learning set are chosen, such that the locations are uniformly distributed across the length of the set.

The locations are assumed to be uniformly distributed in order to use the simplest model for even sampling across varying regions. It is also a reasonable assumption if the underlying system is unknown.

In order to examine the sensitivity of the combined forecast to the uncertainty in the initial condition *perturbed*, noisy initial conditions are constructed. Specifically, a *perturbed initial condition* η_i is one of the realisations of a random variable η, such that $\eta \sim N(\mu_\eta, \sigma_\eta)$, with μ_η defined by a value of a point on a learning set $\mu_\eta = s_t$ for a chosen time t. The variance σ_η is chosen

to be the same for each location. Then a set of perturbed initial conditions at each location is produced by generating M realisations of η. Again we have chosen to consider normally distributed perturbation. Other distributions, such as Uniform, might also be interesting to consider.

(4) **Generating ensembles using those initial conditions**: Once sets of the initial conditions have been generated they are used as an input into a forecasting model. As we are assuming the perfect model scenario the equations that have generated the learning set are also used as the forecasting model. As a result, L ensemble forecasts are produced (one for each random location on the learning set), extending up to lead time N_e, where each ensemble consists of M ensemble members (one for each perturbation of the initial condition).

(5) **Parameter optimisation using Ignorance.**

We are interested in transforming ensemble forecasts into a probabilistic form, and for this we use the procedure known as *kernel-dressing* [**6**]. This is achieved by assuming that ensemble members, at lead time t, are draws from a probability density function $p_f^t(y)$, which is defined by equation 34.

$$p_f^t(y) = \frac{1}{\sigma_m^t\sqrt{2\pi}} \frac{\left(\sum_{i=1}^{M}\left[\exp\left(\frac{-(y-\mu_i^t)^2}{2(\sigma_m^t)^2}\right)\right]\right)}{M}; \tag{34}$$

The form of $p_f^t(y)$ is derived by fitting a Normal distribution around each ensemble member μ_i^t at a given lead time t, and then averaging

over M normal distributions to get an ensemble based probability distribution. As a result, σ_m^t is the first parameter that has to be determined.

The second unknown parameter is α^t, which determines how much information from the forecasting model is used in the final forecast $F^t(y)$, equation 25. We proceed by simultaneously choosing σ_m^t and α^t by minimising the Ignorance skill score [**45**] of $F^t(y)$, measured in bits. In our case the L verification sets are subsets of the learning set of length $N_e - 1$ that start one day after the L initial dates t. Each location has an ensemble forecast and it's corresponding verification set. Hence at each location, for every given lead time t, the Ignorance surface of $F^t(y)$ is defined by:

$$\left\langle IGN_l\left(F^t(y)\right)\right\rangle = -\frac{1}{T}\sum_{j=1}^{T}\log_2\left(\alpha^t\frac{1}{\sigma_m^t\sqrt{2\pi}M}\right.$$
$$\left.\left(\sum_{i=1}^{M}\left[\exp\left(\frac{-\left(y_j-\mu_i^t\right)^2}{2\left(\sigma_m^t\right)^2}\right)\right]\right)+\left(1-\alpha^t\right)p_c\left(y_j\right)\right)\Bigg]\Bigg). \tag{35}$$

The Ignorance skill score is computed for every location, and then an average of those scores is taken. The parameters α^t and σ_m^t, for every given lead time t, are the values that minimise the average Ignorance surface of $F^t(y)$:

$$\min_{\alpha^t,\sigma_m^t}\left(\frac{1}{L}\sum_{l=1}^{L}\left[IGN_l\left(F^t(y)\right)\right]\right). \tag{36}$$

The minimum on the surface is found using the Conjugate Gradient minimisation algorithm [47]. The Conjugate Gradient minimisation will search for $\alpha^t \in (-\infty, \infty)$ and $\sigma_m^t \in (-\infty, \infty)$, however, we require $0 \leq \alpha^t \leq 1$ and $\sigma_m^t \geq 0$. in order to overcome this problem the *Logit* parameter transformation is used, and new parameters α^{*t}, equation 29, and σ_m^{*t}, equation 30, are introduced.

The Logit transformation allows the Conjugate Gradient method to freely search for the parameter values of α^{*t} and σ_m^{*t} in the range from $-\infty$ to ∞. Using equations 29 and 30, α^t and σ_m^t can then be written as:

$$\alpha^t = \frac{\exp(\alpha^{*t})}{\exp(\alpha^{*t}) + 1} \tag{37}$$

$$\sigma_m^t = \exp(\sigma_m^{*t}). \tag{38}$$

Then substituting for α^t and σ_m^t as defined in equations 37, 38 into equation 36 and applying averaging over locations at a fixed time we obtain the average ignorance, equation 39. Now equation 39 can be minimised by the Conjugate Gradient minimisation algorithm.

$$IGN\left(F^t\left(y\right)\right) = \frac{1}{L}\left(\sum_L\left[-\frac{1}{T}\sum_{j=1}^{T}\log_2\left(\frac{\exp(\alpha^{*t})}{\exp(\alpha^{*t})+1}\right.\right.\right.$$
$$\left.\left.\left.\left(\frac{1}{\exp(\sigma_m^{*t})\sqrt{2\pi M}}\sum_{i=1}^{M}\left[\exp\left(\frac{-\left(y_j-\mu_i^t\right)^2}{2(\exp(\sigma_m^{*t}))^2}\right)\right]\right)+\right.\right.\right.$$
$$\left.\left.\left.+\left(1-\frac{\exp(\alpha^{*t})}{\exp(\alpha^{*t})+1}\right)p_c\left(y_j\right)\right)\right]\right).$$

(39)

Finally, the Conjugate Gradient method requires the first partial derivatives of a function with respect to the parameters that need to be optimised. $\frac{\partial IGN\left(F^t(y)\right)}{\partial\alpha^{*t}}$ and $\frac{\partial IGN\left(F^t(y)\right)}{\partial\sigma_m^{*t}(t)}$ are given in equations 40 and 41 respectively.

$$\frac{\partial IGN\left(F^t(y)\right)}{\partial\alpha^{*t}} = \frac{1}{L}\left(\sum_L\left[-\frac{1}{T}\sum_{j=1}^{T}\log_2\left(\frac{\exp(\alpha^{*t})}{(\exp(\alpha^{*t})+1)^2}\right.\right.\right.$$
$$\left.\left.\left.\left(\frac{1}{\exp(\sigma_m^{*t})\sqrt{2\pi M}}\sum_{i=1}^{M}\left[\exp\left(\frac{-\left(y_j-\mu_i^t\right)^2}{2(\exp(\sigma_m^{*t}))^2}\right)\right]\right)+\right.\right.\right.$$
$$\left.\left.\left.+\left(1-\frac{\exp(\alpha^{*t})}{(\exp(\alpha^{*t})+1)^2}\right)p_c\left(y_j\right)\right)\right]\right).$$

(40)

$$\frac{\partial IGN\left(F^t(y)\right)}{\partial\sigma_m^{*t}} = \frac{1}{L}\left(\sum_L\left[-\frac{1}{T}\sum_{j=1}^{T}\log_2\left(\left(\frac{\exp(\alpha^{*t})}{(\exp(\alpha^{*t})+1)^2}\right)\right.\right.\right.$$
$$\left.\left.\left.\left(\frac{1}{\exp(\sigma_m^{*t})\sqrt{2\pi M}}\right)\left(\sum_{i=1}^{M}\left[\exp\left(\frac{-\left(y_j-\mu_i^t\right)^2}{2(\exp(\sigma_m^{*t}))^2}\right)\left(\frac{\left(y_j-\mu_i^t\right)^2}{(\exp(\sigma_m^{*t}))^2}-1\right)\right]\right)+\right.\right.\right.$$
$$\left.\left.\left.+\left(1-\frac{\exp(\alpha^{*t})}{(\exp(\alpha^{*t})+1)^2}\right)p_c\left(y_j\right)\right)\right]\right).$$

(41)

Once the optimal α^{*t} and σ_m^{*t} have been determined one can compute corresponding values of α^t and σ_m^t using equations 29 and 30 respectively.

(6) **Issuing the out of sample forecast**

The final stage of the process combines the results obtained in all the previous steps. A first a point of the system being studied, denoted for convenience as s_t, is chosen out of sample. This point defines the start of the prediction period and is used to construct perturbed initial conditions, as described in step 3.

The next step involves determining $p_f^t(y)$ by kernel-dressing the forecast ensemble for each lead time t using equation 34 and the σ_m^t calculated in step 5. Finally the out of sample forecast is given by the equation 25 with the $p_c(y)$ constructed in step 2 and α^t also computed in step 5.

3. Lorenz Experiment

The method is first tested on a chaotic system. Chaotic systems exhibit sensitivity to the initial conditions, which means that only initial uncertainty results in a forecast uncertainty that grows rapidly with time [**27**]. The same characteristic is also observed in weather forecasts (see chapter 2), although this is due not only to the uncertainty in initial conditions, but also operational and model error. As a result, this property of chaotic systems allows us to conduct a controlled experiment using a method that one wants to apply to weather forecasts, but in a much simpler setting.

a	r	b	x_0	y_0	z_0
10	28	$\frac{8}{3}$	0	-0.01	9

TABLE 2. Parameters and initial conditions of the Lorenz system for the chaotic state.

The experiment was first carried out on a data set generated using the Lorenz 63 dynamical system [**44**]. The Lorenz system is nonlinear, three-dimensional and deterministic. The equations that describe the system are presented in equation 42 and were derived from the convection arising in the equations of the atmosphere. For some parameter values [**46**] the system is chaotic. Such parameters were chosen and are shown in table 2 together with the initial conditions.

$$\begin{aligned} \frac{\partial x}{\partial t} &= a(y-x), \\ \frac{\partial y}{\partial t} &= -xz + rx - y, \\ \frac{\partial z}{\partial t} &= xy - bz. \end{aligned}$$

(42)

In order to generate the Lorenz data a numerical integration scheme has to be used. The classical fourth-order RungeKutta method [**47**] has been chosen and extended to the three dimensional system (see section 2.1).

Now everything is set in order to conduct the experiment and produce out of sample combination forecasts for noisy Lorenz data (particularly the

N_{ls}	M	T_e	L	N_{BINS}	σ_ε	σ_η	T_{MAX}	h
72 L.s	100	7.3 L.s	200	128	0.2	0.4	7.3 L.s	0.01

TABLE 3. Parameter values used in the production of combination forecasts for the Lorenz data experiment, where L.s. stands for Lorenz seconds.

x component). The summary of parameter values used in production of combination forecasts in the Lorenz experiment are presented in table 3.

3.1. Generating Lorenz learning set and climatology. First Runge-Kutta numerical integration (equations 32 and 33) with $h = 0.01$ are used to generate $N_{ls} = 72$ Lorenz seconds of Lorenz data, which forms the Lorenz attractor [**27**] presented in figure 1. It can be seen from this picture that some parts of the Lorenz attractor are very flat (area of the two 'wings'). It is always a benefit to examine the system visually, as in this case, if we choose a point on the attractor that falls into the flat area (which we do later on in this chapter), when that point is perturbed with nose, the 'noise ball' around it will occupy all parts of the space, in particular parts of the space where the Lorenz observations will never occupy. We do not study in this thesis the effects of perturbing initial conditions such that they occupy part of the space that is not incorporated by the attractor of that chosen system. It is something, however, to bare in mind for the further research of the noise propagation through the equations of motion of a chosen chaotic system.

In this experiment, to minimise computer time and for simplicity, only the x component of the Lorenz data is used, however the method can be

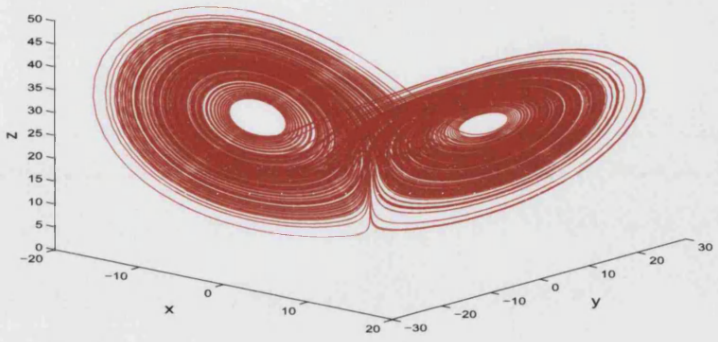


FIGURE 1. Lorenz attractor.

extended to multiple dimensions. Once the data has been generated, a Normally distributed (equation (26)) nose $\varepsilon \sim N(\mu_\varepsilon, \sigma_\varepsilon)$ is added. The value of $\sigma_\varepsilon = 0.2$ is chosen such that generated realisations of ε range from $(-1, 1)$, which is approximately 2% of the size of the Lorenz attractor in the x component direction. The realisations of ε_i are generated using Matlab's randn function [**56**]. A fragment of the resulting noisy x component is presented in figure 2. The values of the x component lie in the ranges $(-20\ 17)$ and exhibit growing oscillations that are followed by rapid jumps. Growing oscillations happen when points are moving on a wing of the Lorenz attractor. These are regions with better predictability, compared to the periods where jumps occur in the time series. This corresponds to jumps between the two wings of the attractor.

The next step is to construct the climatology of the learning set, i.e. the climatology of the noisy Lorenz x component. This is achieved by setting the number of bins to be: $N_{BINS} = 128$, which is a reasonable number of bins to be able to distinguish details in the observed relative frequency

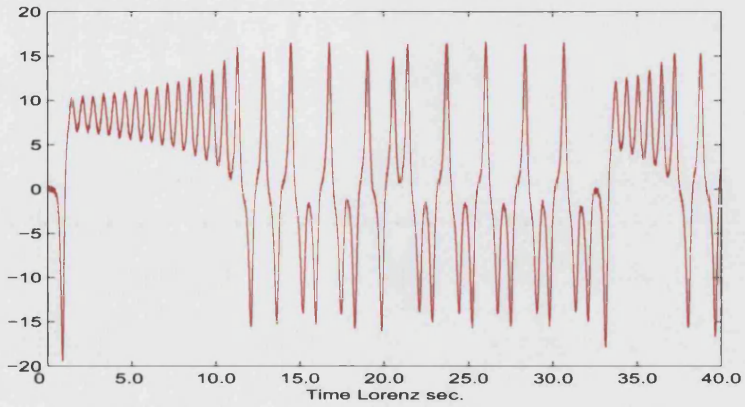


FIGURE 2. Fragment of the x component of the Lorenz dynamical system with additive noise $\varepsilon \sim N(0, 0.2)$.

with regards to number of points in the learning set and the range of data. The bins are equal in size and range from the minimum to the maximum observed values in the learning set.

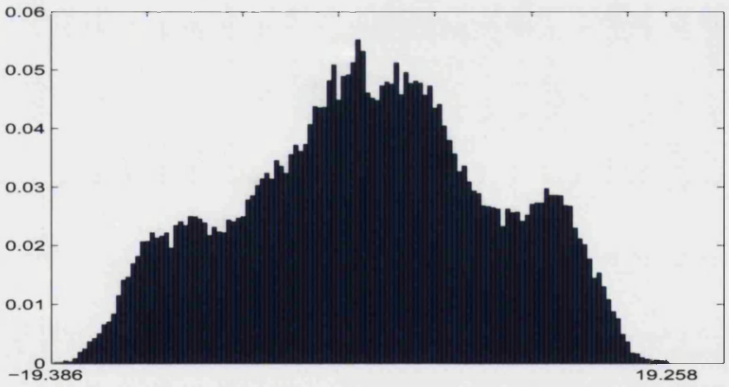


FIGURE 3. Historic relative frequency of the noisy x component of the Lorenz system, constructed using 128 equally sized bins ranging from the minimum to the maximum values of the observed 72 Lorenz sec of data.

3.2. Generating perturbed initial conditions and resulting ensemble forecasts. The length of the learning set N_{ls} defines the upper boundary of the uniform distribution, in other words $a = 1$ and $b = 72$ Lorenz sec, equation (27). Then $L = 200$ random draws from that uniform distribution are taken using Matlab's rand function.

Once the locations on the learning set are determined, sets of perturbed initial conditions are constructed. The value of the Lorenz data at each location s_t determines the mean of the perturbation noise $\eta \sim N(\mu_\eta, \sigma_\eta)$ where $\mu_\eta = s_t$ and $\sigma_\eta = 0.404145$. The chosen value of σ_η produces noise that ranges from $(-1.9\ 1.9)$ from s_t and is approximately 4.5 % size of the attractor in the x component direction. We want to examine the consistency and characteristics of a combinational forecast, for relatively low levels of perturbation noise. Again, the realisations of η_i are generated using Matlab's randn function. As mentioned in the previous section the generated noise will occupy the part of space (for the flat regions of the attractor) that are not occupied by the Lorenz system, see figure 4.

Initial conditions are used to generate ensembles (ensemble forecasts) at each location. Each ensemble is chosen to consist of $M = 100$ members and each ensemble extends up to $T_e = 7.3$ Lorenz seconds. As discussed, the uncertainty in ensemble forecasts will vary for different locations for a given level of noise in initial conditions, as some regions of the Lorenz attractor display better predictability compared to others. This can be observed in figure 5 where ensemble forecasts for three locations are displayed.

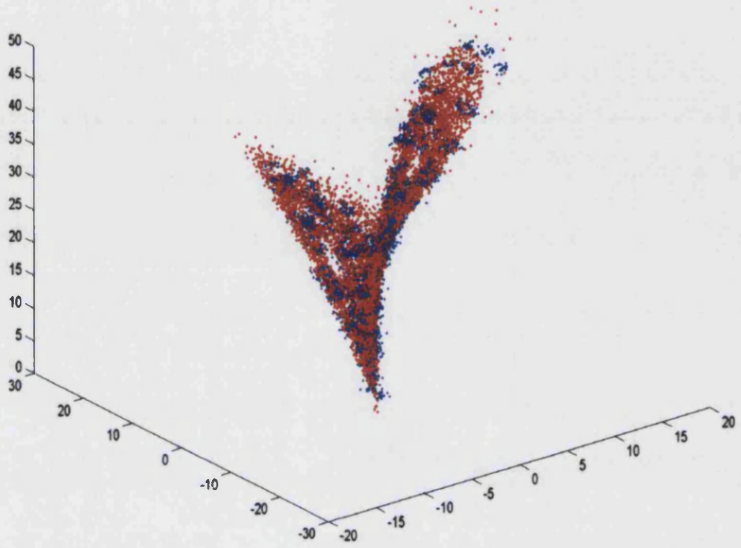


FIGURE 4. Generated initial conditions that has been perturbed with Normally distributed noise with standard deviation 0.404145.

It can be seen that at the first location (top plot) the ensemble forecast is the least certain. The split of the ensemble members happens at approximately 0.1 Lorenz sec., where some ensemble members go to another wing of the Lorenz attractor compared to the verification set. This location corresponds to a less predictable region of the Lorenz attractor, where a small change in an initial condition leads to a trajectory ending up on the opposite wing. At a location displayed in the second plot, ensembles are the most stable out of the three locations displayed, and ensemble members stay close together for up to 1.0 Lorenz sec. This ensemble was generated from initial conditions in a more stable region of the attractor. After one sec. more

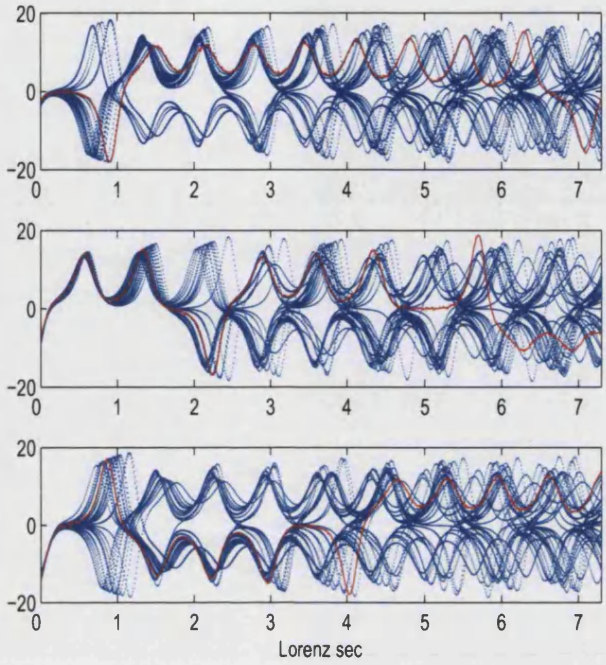


FIGURE 5. Examples of generated ensembles from perturbed initial conditions in three different locations (blue), together with the verification data (red) for the Lorenz experiment.

uncertainty is present and that is evident by the increasing spread[3] of the ensemble members. The split occurs later, at around 1.6 Lorenz sec.

3.3. Kernel dressing: constructing parameter surface using ignorance skill score and finding the minima. The next step in producing combined forecasts, involves parameter estimation: the kernel dressing parameter σ_m^t and the blending parameter α^t. Particularly, we use equation 39 with $T = 1$ L.s, $M = 100$ and μ_i^t- the value of i^{th} ensemble member with $i = 1 \ldots M$ for each initial location, for a given lead time t. y_j is the value

[3]The ensemble spread is defined to be the difference between the max and the min values of the ensemble members at a given time.

of the learning set one step after the perturbation point at each location, i.e s_{t+1}. Equations 40 and 41 are also constructed with the same parameter values, and are provided to the Conjugate Gradient minimisation algorithm. As a result optimal α^{*t} and σ_m^{*t} are found. Finally equations 37 and 38 are used to determine α^t and σ_m^t respectively. Values of α^t and σ_m^t as functions of lead time are presented in figures 6 and 7 respectively.

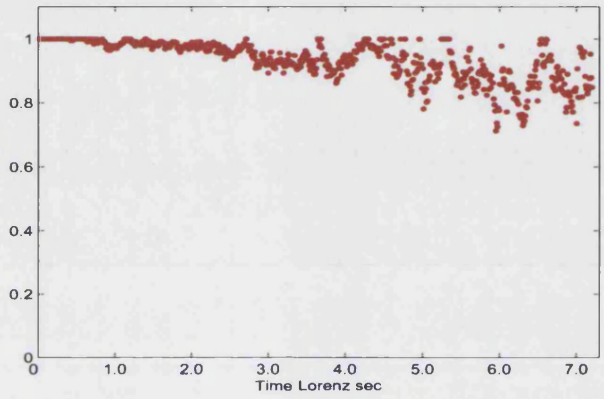


FIGURE 6. Blending parameter for the Lorenz experiment. Red- is the value of the blending parameter α^t with lead time (x- axis). y axis indicate the size of the attractor for the x component of Lorenz system.

The resulting blending parameter α^t remains at 1 for up to 1.0 Lorenz sec., which means that up to 1.0 Lorenz sec. all the information in the combinational forecast is going to come from the ensemble forecast. And the uncertainty exhibited in ensembles will be reflected in the values of the kernel dressing parameter σ_m^t. It is also evident that as lead time increases the values of α^t exhibit a decreasing trend. The instability, however, in the values of α^t increases with lead time.

The values of α^t for the Lorenz experiment were calculated for the maximum lead time of $T_{MAX} = 7.2$ Lorenz sec. The value of α^t only drops to approximately 0.8 at the maximum observed lead time, which means that even after 7.0 Lorenz sec. most of the information (80 %) is taken from the ensemble forecast rather than climatology. This is reasonable, as α^t reflects how well the verification data is captured by the ensemble forecast. By 7.0 Lorenz sec., verification sets are mostly contained within some of the members of the corresponding ensemble forecasts. This therefore results in a larger values of α^t. As lead time increases the verification data falls out of the ensemble more frequently, which courses the decreasing trend in the values of α^t.

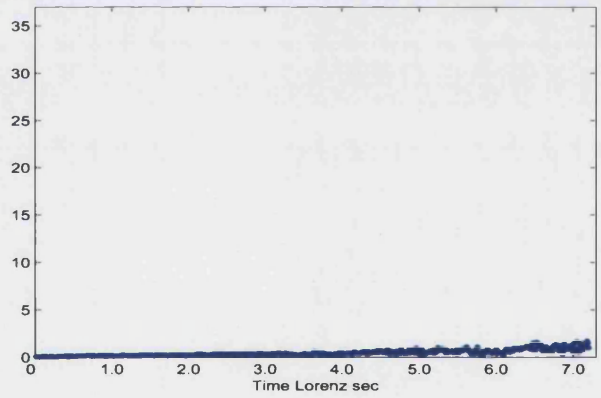


FIGURE 7. The width of the fitted distribution parameter σ_m^t with lead time (x- axis). y axis indicate the size of the attractor for the x component of Lorenz system.

The value of σ_m^t is initially very small compared to the size of the attractor, but it is consistent with the size of the perturbation noise. As α^t is 1 at the initial time all the uncertainty in the combinational forecast has to be

accounted for by σ_m^t. Hence the value of σ_m^t in the initial period should be comparable to the average size of the perturbation noise. As the lead time increases, σ_m^t also increases, however the level of σ_m^t even after $T_{MAX} = 7.2$ is minute compared to the size of the attractor. σ_m^t reflects the average distance of a verification data from members of the corresponding ensemble forecast. Small levels of σ_m^t after 7.2 Lorenz sec. lead to the conclusion that most of the ensemble members at most locations still remain close to the verification set.

3.4. Exploring the influence of noise in the initial conditions and the size of the ensemble on the parameter estimation. In order to examine the sensitivity of the parameter estimation and hence the change in predictability of the combination forecasts to the levels of uncertainty in initial conditions, the experiment was repeated with a higher level of noise. Particularly, the level of noise in perturbed initial conditions has been doubled , particularly $\sigma_\eta = 0.808290$. Figure 8 illustrates the comparison between perturbed initial conditions generated with $\sigma_\eta = 0.404145$ and $\sigma_\eta = 0.808290$.

The change in the level of noise in the initial conditions affects the behaviour of the resulting ensemble forecasts, in particular the ensemble spread. Figure 9 compares ensemble forecasts for the same three locations, displayed in figure 5, produced using two levels of noise in perturbed initial conditions. It can be seen on that picture particularly clearly that the generated noise (for both cases) occupies the part of space where the Lorenz

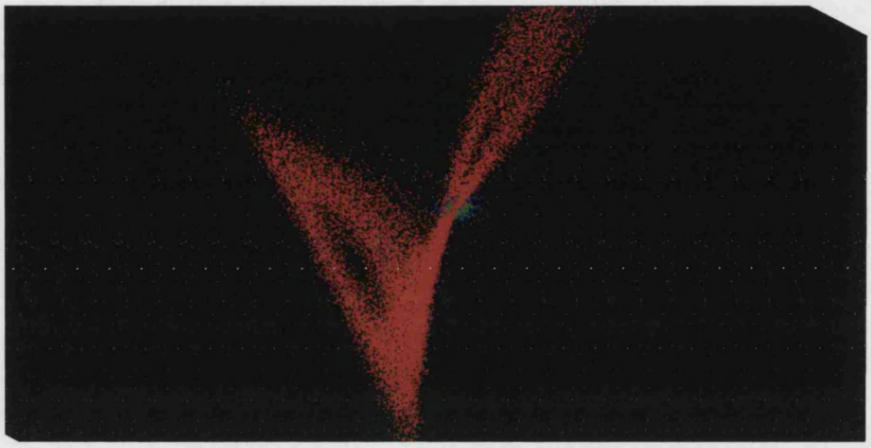


FIGURE 8. Comparison of the perturbed initial conditions at a location on the attractor, where different level of noise has been added. Green- initial conditions constructed with Normally distributed noise with $\sigma_\eta = 0.404145$. Blue are initial conditions constructed with normally distributed noise with standard deviation $\sigma_\eta = 0.808290$. Red-noisy Lorenz attractor.

attractor is not present. Again this is not studied here, but should be investigated. This figure also clearly illustrates the size of the generated noise for two cases and the relative size of the two nosy initial conditions sets with respect to the Lorenz attractor.

It can be observed that ensembles generated using initial conditions with $\sigma_\eta = 0.808290$, generally exhibit larger spread. Also the split in the ensemble members, where some of the trajectories travel to the opposite wing happens earlier, as expected. In other words increased level of the perturbation noise leads to greater uncertainty in the ensemble forecast. As a result of the larger level of the perturbation noise, ensembles contain less precise

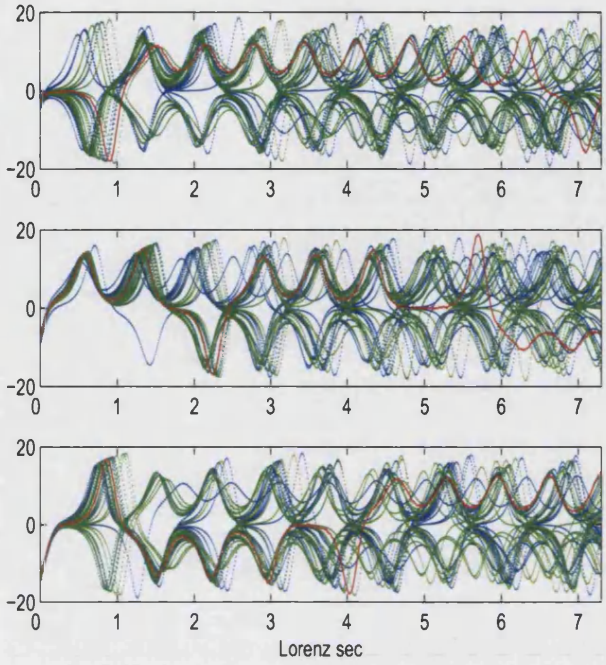


FIGURE 9. Comparison of ensembles generated using different levels of noise at a given location. Green- ensemble generated using normally perturbed initial conditions with $\sigma_\eta = 0.404145$; Blue- ensembles initialised with normally distributed initial conditions with doubled level of $\sigma_\eta = 0.808290$ and red is the verification data.

information about the combined forecast. This is evident in the values of α^t and σ_m^t. The values of α^t and σ_m^t, as functions of lead time t, calculated using perturbed initial conditions with the doubled level of noise are compared to the previously computed α^t and σ_m^t, and the results shown in figures 10 and 11 respectively.

For larger levels of noise α^t stays at exactly 1.0 for a shorter period of time and drops to a lower level as lead time increases. By 7.2 Lorenz

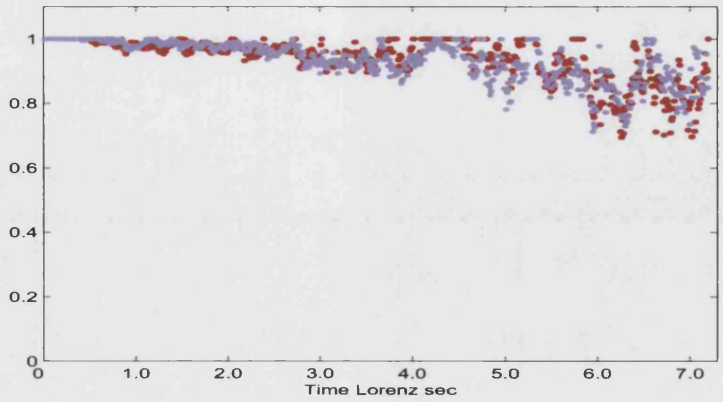


FIGURE 10. α^t for increasing lead times computed as the result of different levels of noise in the initial conditions. Magenta-blending parameter α^t for the smaller noise level. Red is α^t for the doubled level of noise.

sec. however, the level of α^t still remains at a large level of approximately 0.7. The difference between the combined forecast parameters (α^t and σ_m^t) for those two different levels of noise is small, as even the doubled level of perturbation noise is still small compared to the size of the attractor in the x component direction.

Again, larger levels of noise in the perturbed initial conditions result in a larger σ_m^t, on average. If the perturbation noise is very large, say 30 % the size of the attractor in the x component direction, the values of σ_m^t also start at a level of 30 % of the size of the attractor. α^t in this case will remain at 1.0 for a long time as the ensemble will always capture the corresponding learning set. If we, however, start with the perturbation noise smaller than the observation noise, the determined α^t will initially be even less than 1.0 and will also reduce further in value as lead time increases.

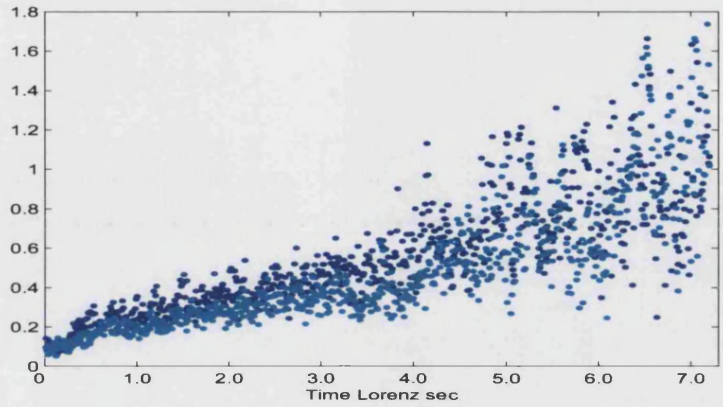


FIGURE 11.]

Zoom into the σ_m^t values for increasing lead times computed as the result of different levels of noise in the initial conditions. Cyan- s.d. of the dressing kernel - σ_m^t for the smaller noise level. Blue is σ_m^t for the doubled level of noise respectively.

We are also interested in sensitivity of the combined forecast's calibrated parameters to the ensemble size M. Therefore we examine what happens to the estimates of the parameters when the ensemble size is reduced: specifically $M = 50$, $M = 25$ and $M = 12$ have been used. Figure 12 and 13 illustrate changes observed in α^t and σ_m^t respectively that occur when the ensemble size is reduced.

It can be noted that the general trends observed in the parameter values with $M = 100$ remain the same. Particularly, as lead time increases the values of σ_m^t grow and the values of α^t exhibit a decreasing trend. Additionally it can be seen that the values of σ_m^t increase as the size of the ensemble decreases, which is expected. Previously it has been observed in ensemble forecasts (for $M = 100$), that most of the ensemble members stay close to

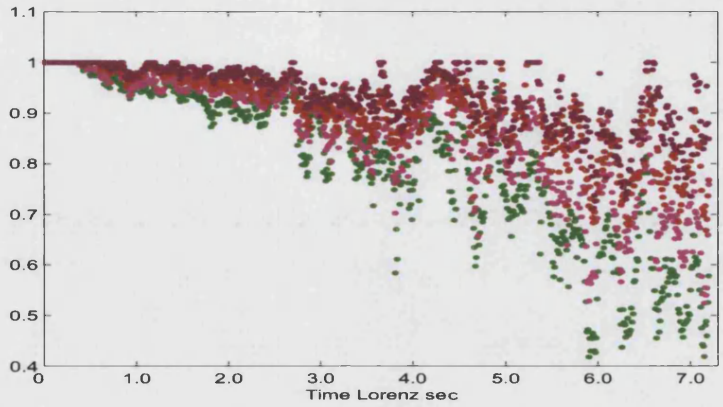


FIGURE 12. Contrasting α^t resulted from different ensemble sizes, colours darken with the increase in ensemble size: The green is α_m^t respectively generated using ensemble consisting of 12 members; magenta is α^t of the ensemble composed of 25 members; red is α^t for the 50 member ensemble; and finally deep burgundy is α^t that has been generated for the full ensemble size (100 members).

the verification set particulary for smaller lead times. As a result, as the number of ensemble members reduces, members that are far away from the verification set will have more influence over the average distance between the ensemble members and the corresponding verification set. The values of α^t decreases as the size of the ensemble is reduced, which is expected as the verification set is more likely to fall outside the boundaries of the corresponding ensemble.

3.5. Issuing forecasts. The final stage of the production of the combined forecast of a time series uses results from all previous parts of the experiment. First we start with a point of the Lorenz time series that one

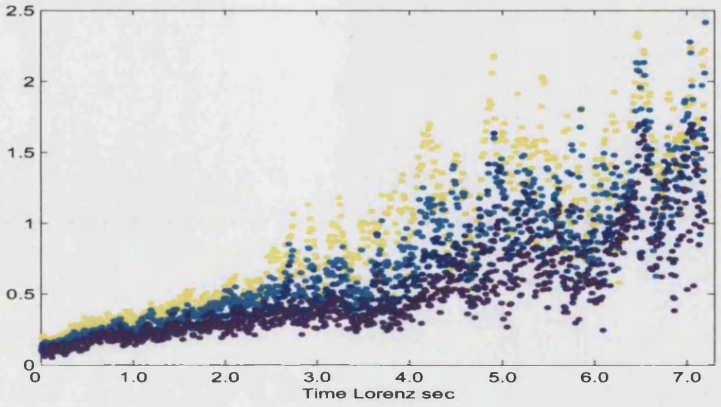


FIGURE 13. Contrasting σ_m^t resulted from different ensemble sizes, colours darken with the increase in ensemble size: Yellow - σ_m^t generated using ensemble consisting of 12 members; Cyan is σ_m^t of the ensemble composed of 25 members; Blue is σ_m^t for the 50 member ensemble; and finally Dark purple is σ_m^t that has been generated for the full ensemble size (100 members).

observes and beyond which one wishes to forecast. This data point is used to generate perturbed initial conditions with the same level of perturbation noise, i.e. $\sigma_\eta = 0.404145$, that was used in the parameter estimation. Finally, the ensemble forecast is produced using these perturbed initial conditions, see figure 14.

Red- is the out of sample Lorenz x component time series that one wishes to predict in the form of a probability distribution. So for the chosen parameters of the Lorenz experiment: $\sigma_\eta = 0.404145$, $M = 100$, $T_{MAX}, T_e =$ 7.2 Lorenz sec.; the previously constructed climatology (figure 42); and the calibrated α^t and σ_m^t (figures 6 and 7); the combinational forecast in the

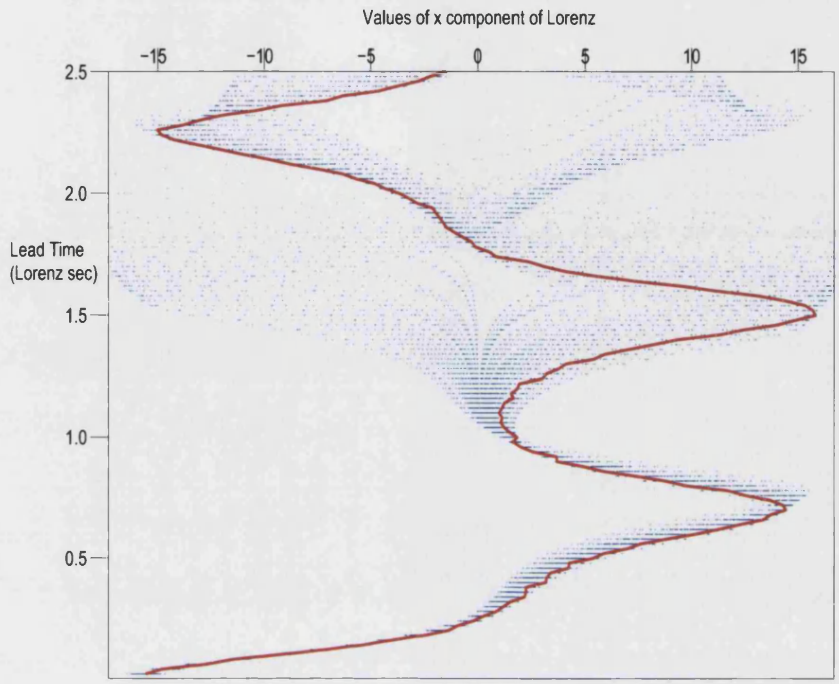


FIGURE 14. Out of sample ensemble forecast for the Lorenz x component that is used in the final probability forecast (blue) and the verification (red).

form of probability function is produced, using equations 25 and 34. Figure 15 shows combinational forecast distributions for up to lead time $t = 2.5$ Lorenz sec.

The combinational forecast obtained displays higher certainty up to 0.5 Lorenz sec. After that the spread in the forecasting distribution increases, providing a less precise forecast. For lead times around 1.0 Lorenz sec. the distributions become tighter for a short period, and afterwards widen where we observe a dip in the levels of α^t, which means that the climatology gets introduced into the combinational forecast. It can be seen that as lead time increases the spread in the forecasting distribution increases. The combined

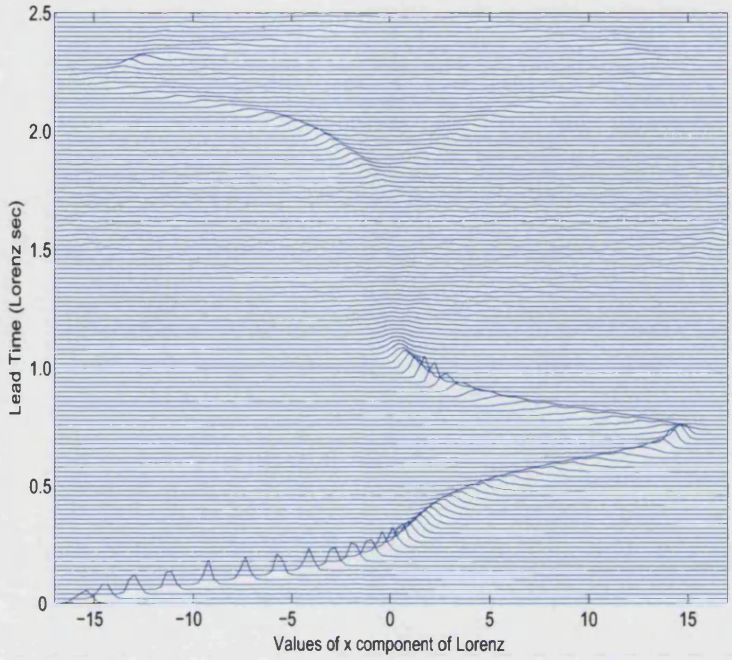


FIGURE 15. Final probability forecast issued for the Lorenz x component using combinational forecasts.

forecast produced closely reflects the behaviour of the ensemble forecast, figure 14. This is reasonable, as the values of α^t remain close to 1.0 for the forecasting period.

T	R	x_0	y_0	z_0
6	20	0.1	0	0

TABLE 4. Parameters and initial conditions of the Moore-Spiegel system.

N_{ls}	M	T_e	L	N_{BINS}	σ_ε	σ_η	T_{MAX}	h
73 M-S.s [4]	100	110 M-S.s	200	128	0.02	0.057735	110 M-S.s	0.01

TABLE 5. Parameters of Moor-Spiegel experiment

4. Perfect model scenario: Moore-Spiegel experiment.

The method of combined forecasts has been also tested on another chaotic system: Moore-Spiegel [**27**]. The Moore-Spiegel dynamical system describes the evolution of gas flows in the centre of a galaxy. The equations that describe the system are given in equation 43.

$$\begin{aligned} \frac{dx}{dt} &= y, \\ \frac{dy}{dt} &= z, \\ \frac{dz}{dt} &= -z - \left(T - R + Rx^2\right) y - Tx. \end{aligned} \tag{43}$$

For some parameters the system exhibits chaotic behaviour. These parameter values have been chosen for this experiment and are presented in table 4 together with the initial conditions.

The parameters that are used to produce combined forecasts are summarised in table 5.

4.1. Generation of the learning set and climatology for the Moore-Spiegel chaotic oscillator. In order to generate Moore-Spiegel data Runge-Kutta numerical integration is also used, equations 32 and 33 with $h = 0.01$. The length of the leaning set was chosen to be $N_{ls} = 73$ Moore-Spiegel seconds. Additionally additive noise ε is generated where $\varepsilon \sim N(\mu_\varepsilon, \sigma_\varepsilon)$ with $\mu_\varepsilon = 0$ and $\sigma_\varepsilon = 0.02$. The realisations of ε_t are generated using Matlab's randn function. The resulting noisy attractor for the Moore-Spiegel system, is illustrated in figure 16. Again visual illustration of the attractor allows to examine the parts of the space that the system occupies.

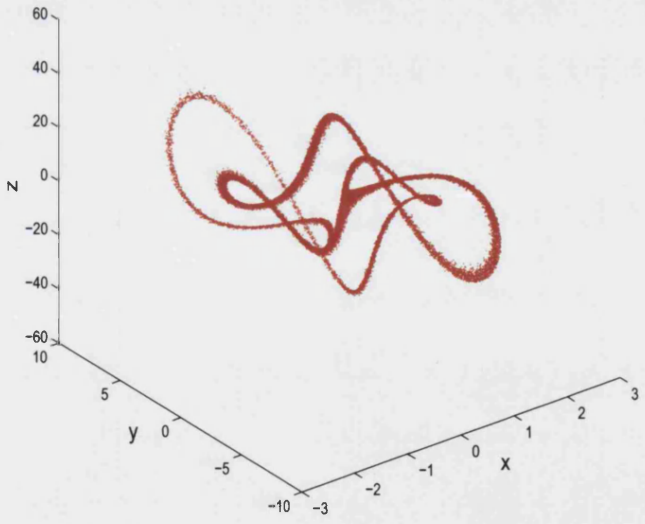


FIGURE 16. Moore-Spiegel attractor with added noise, where added noise is normally distributed with zero mean and s.d. = 0.02.

Again the learning set will only comprise of the x component of the Moore-Spiegel data. The time series corresponding to the x component of the Moore-Spiegel system is illustrated on figure 17.

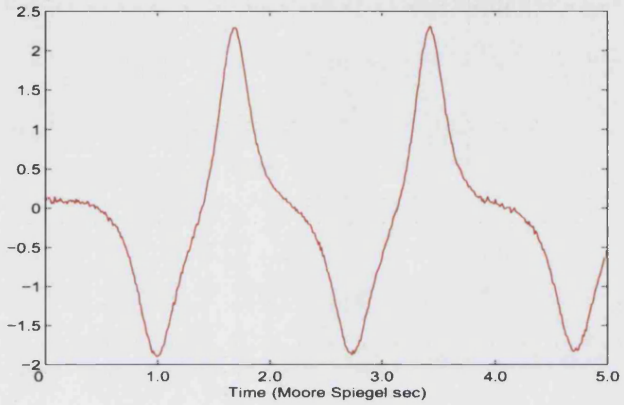


FIGURE 17. The noisy x component of the Moore-Spiegel dynamic isolator.

The climatology of the learning set is constructed using $N_{BINS} = 128$ bins. The size of the bins is identical and the range is determined by the minimum and the maximum observed values in the learning set. The climatology is displayed in figure 18.

4.2. Construction of initial conditions and ensemble forecasts. The first stage in construction of perturbed initial conditions involves determination of locations on the attractor where these conditions will be constructed. Again, the locations are chosen to be uniformly distributed, with $b = 73$ Moore-Spiegel sec is chosen to be used in equation 27. $L = 200$ random draws from that uniform distribution are generated using Matlab's rand function.

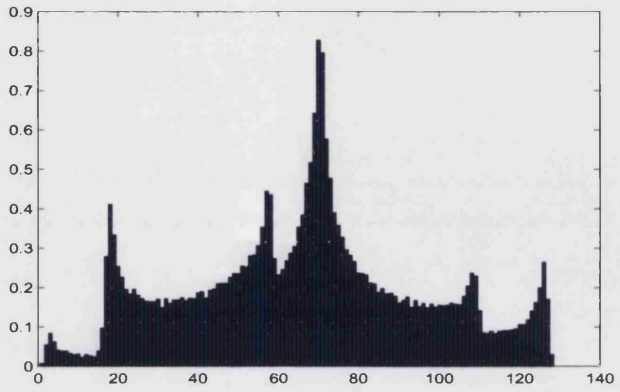


FIGURE 18. Climatology of the generated noisy x component of the Moore-Spiegel system, where 73 Moore-Spiegel seconds of data has been used distributed between equally sized 128 bins, ranging from the maximum and minimum vales observed in the data.

At each chosen location the value of the Moore-Spiegel s_t determines the mean of the perturbation noise $\mu_\eta = s_t$, with $\eta \sim N(\mu_\eta, \sigma_\eta)$. The standard deviation of the perturbation noise is set to be $\sigma_\eta = 0.057735$, which produces noise in the range $(-0.1, 0.1)$ from s_t and is approximately 4% size of the attractor in the x component direction. This Perturbation noise is generated using Matlab's randn function. The perturbed initial conditions are displayed on figure 19 with relation to the attractor space. We use figure 19 to make sure that the perturbed initial conditions are distributed 'reasonably uniformly' on the attractor. And actually in this case it can be seen in figure 19 that there are some regions of the attractor that are not sampled by the chosen initial conditions.

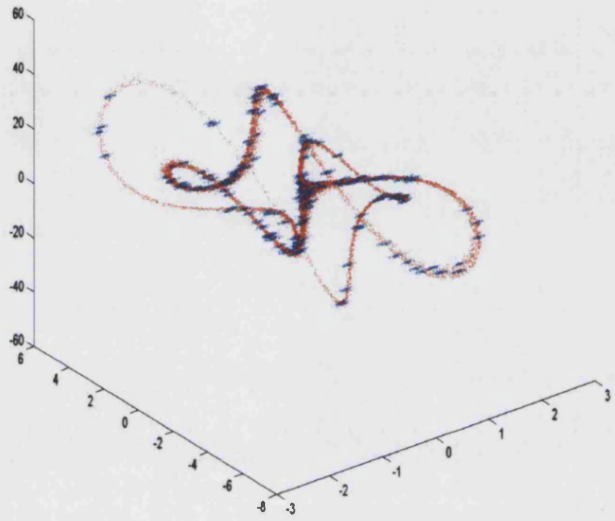


FIGURE 19. Normally distributed perturbed 200 initial conditions with s.d. $= 0.057735$.

Next ensemble forecasts are generated using the equations that describe the Moore-Spiegel system, equation 43. Each ensemble is chosen to consist of $M = 100$ members and each ensemble extends up to $T_e = 110$ Moore-Spiegel sec. Three examples of ensemble forecasts generated together with the verification in three locations are shown on figure 20.

As in the Lorenz experiment, in the Moor-Spiegel experiment different locations exhibit different predictability, and this is evident in the ensemble forecasts, particularly ensemble spread. It can be observed that the location corresponding to the bottom plot is the most stable, and the ensemble forecast produced has very little uncertainty, and most of the ensemble members are close to to the verification set for up to 6 Moore-Spiegel sec. After which

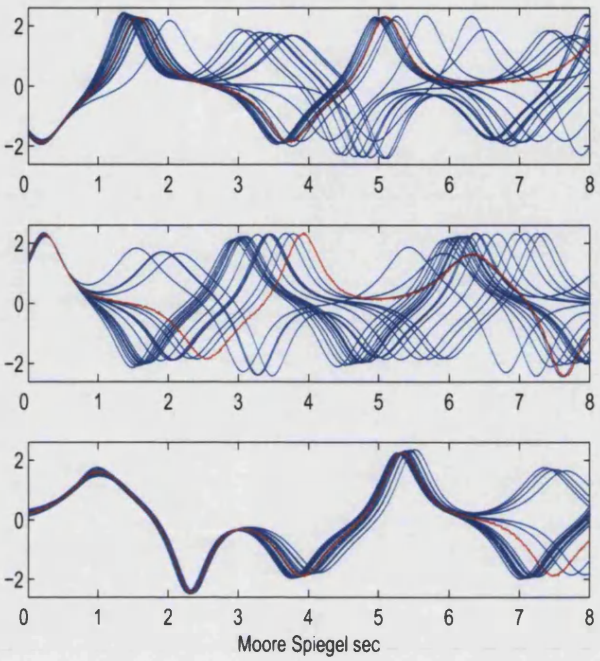


FIGURE 20. Examples of ensembles generated using perturbed initial conditions at a given locations: blue are the ensembles and red are corresponding verification sets.

a split in the ensemble forecast occurs. In comparison the middle plot shows on ensemble which splits before 1.0 Moore-Spiegel sec. and is much more volatile compared to the others.

4.3. Kernel dressing of re-sampled Moore-Spiegel ensembles with minimisation of ignorance skill score based parameter surface. We proceed by estimating the values of α^t and σ_m^t. We are particularly interested in what happens for extended lead times. For the purpose of saving computation time, ensembles and the corresponding verification sets are re-sampled, and every 10th point is taken. Equations 39, 40 and 41 are used in the Conjugate Gradient minimisation algorithm to estimate the

kernel dressing parameter and the blending parameter. Again equations 37 and 38 are used to determine α^t and σ_m^t respectively after the optimisation is performed. Values of α^t and σ_m^t as functions of lead time are presented in figures 21 and 22.

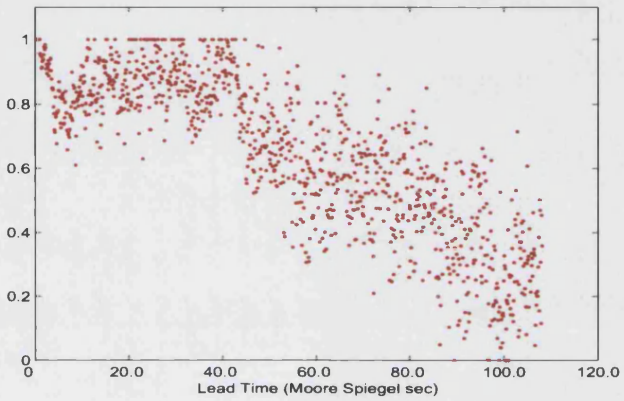


FIGURE 21. Values of parameter α_t for the re-sampled x component of the Moore-Spiegel system with lead time.

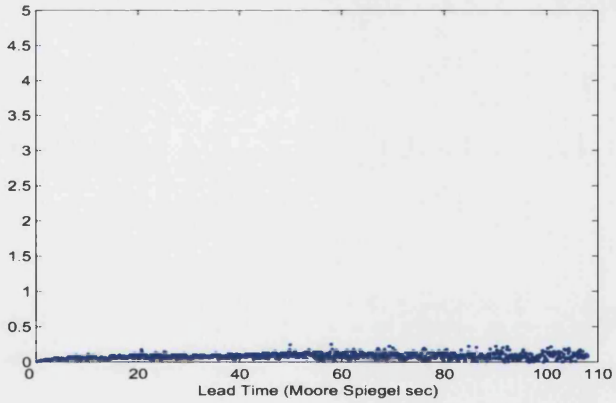


FIGURE 22. Values of parameter σ_t for the re-sampled x component of the Moore-Spiegel system with lead time.

It can be seen, that as in the Lorenz experiment the values of α^t decrease with time and the values of σ_m^t increase with time. Even after a long lead

time $T_{MAX} = 110$ Moor-Spiegel sec. the value of σ_m^t still remains very small compared to the values of α^t, which reduce dramatically as lead time increases.

It can be concluded, after these experiments, that for the observed chaotic systems the values of σ_m^t starts at the level of the perturbation noise used to generate initial conditions. With the increase, in the lead time σ_m^t increases but only by a very small magnitude. The values of the blending parameter α^t exhibit much more dramatic changes in value as the lead time increases. At the initial time α^t starts at 1.0, unless the perturbation noise added is much smaller than the additive noise. As the lead time increases $\alpha^t \to 0$. The spread in the values of both parameters increases with increased lead time. This is reasonable as with increased lead time more and more ensemble members end up in a different region of space compared to the corresponding verification sets, which leads to increased levels of σ_m^t and decreased levels of α^t. Later ensembles come topgether again, due to the cyclical nature of the chosen systems, which results in the levels of σ_m^t being dramatically reduced and levels of α^t being dramatically increased.

4.4. Producing final probability forecast. Lastly, the combined forecast, in the form of a probability distribution for the Moore-Spiegel x component is produced. First, the out of sample Moore-Spiegel data point determines the mean of the perturbation noise and the same level of standard deviation $\sigma_\eta = 0.057735$ is used to construct the perturbed initial conditions. Then the resulting ensemble forecast is constructed as illustrated in

figure 14. Again the equations of Moore-Spiegel are used to construct the ensemble.

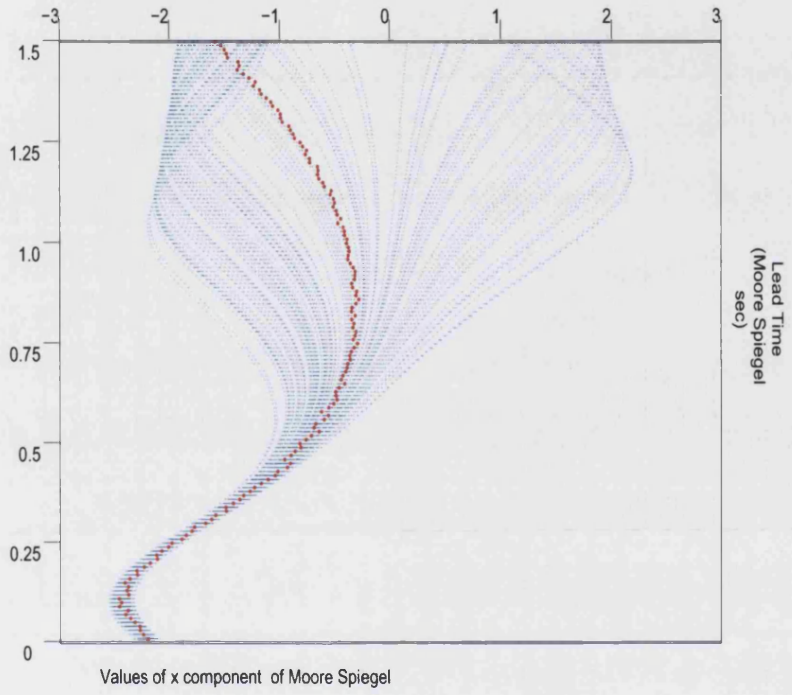


FIGURE 23. Ensemble forecast produced for the x component that is dressed and combined with climatology to produce final probability forecast. Blue- ensemble and red- verification data.

The final forecast of the time series in the Moore-Spiegel experiment with the parameters: $\sigma_\eta = 0.057735$, $M = 100$, $T_{MAX}, T_e = 110$ Moore-Spiegel sec.; the climatology as shown on figure 18; and using the optimal α^t and σ_m^t (figures 21 and 22); is produced, using equations 25 and 34. Figure 24 shows this forecast distributions for up to lead time $t = 1.5$ Moore-Spiegel sec.

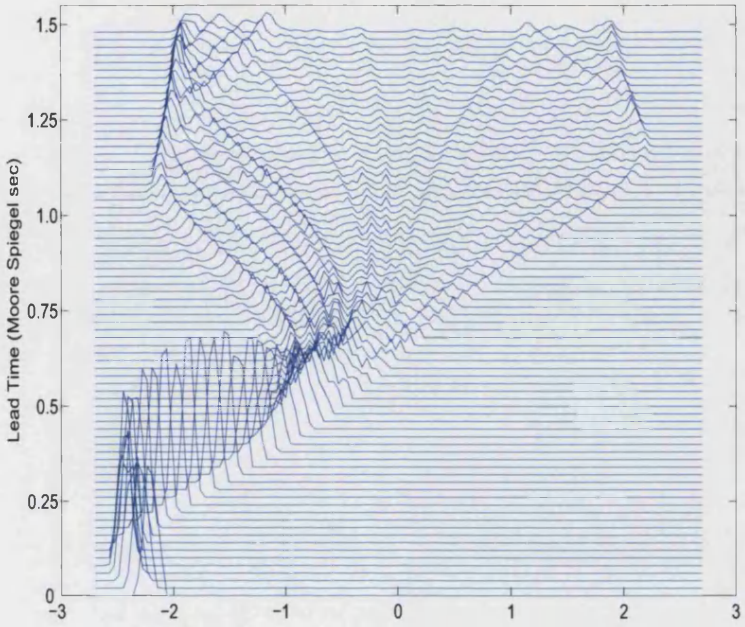


FIGURE 24. Final probability forecast for lead times issued for the x component of the Moore-Spiegel chaotic isolator. x axis indicate the value of the forecasted variable and y axis is lead time.

Again the behaviour of the ensemble forecasts is closely mimicked by the combined probability forecasts produced. Probability forecasts are produced for up to 0.7 Moore-Spiegel sec. with little uncertainty.

5. The perfect model experiment in the statistical modelling framework: Threshold Autoregressive Process

The method for long term forecasting of a time series using combined forecasts was also tested on a data set generated by a classical time series model. Real weather data exhibits jagged and irregular behaviour on short term scales and this test has been performed in order to examine the effectiveness of the method on a time series that is mostly driven by a random component, in contrast to time series generated by a chaotic system.

The purpose is to examine the influence of information produced by a forecasting model in the combined forecast as lead time increases. Additionally, an exact analytical probabilistic forecast can be constructed in this framework, which then can be compared to the numerically computed prediction.

The system to be tested is: A threshold autoregressive process of order three (TAR3) [**54**]. The TAR3 process is defined by equation 44.

$$\begin{aligned} x_{t+1} &= C_1 + a_1 x_t + b_1 x_{t-1} + c_1 x_{t-2} + \xi^1_{t+1} \; if \; x_t < D; \\ x_{t+1} &= C_2 + a_2 x_t + b_2 x_{t-1} + c_2 x_{t-2} + \xi^2_{t+1} \; if \; x_t > D; \end{aligned}$$

(44)

where t is a time step, D is a threshold constant and the random components ξ^1_t and ξ^1_t are chosen to be independent gaussian random variables $\xi^1_t \sim N\left(0, \sigma_{\xi^1_t}\right)$, $\xi^2_t \sim N\left(0, \sigma_{\xi^2_t}\right)$, with $\sigma_{\xi^1_t} = 0.3$, $\sigma_{\xi^2_t} = 0.1$. The list of

a_1	b_1	c_1	a_2	b_2	c_2	C_1	C_2	x_0	y_0	z_0	D	s.d. of ε^1	s.d. of ε^2	Mean of ε^1 and ε^2
0.1	0.5	0.4	0.7	-0.2	-0.7	0	0	0.1	0.5	-0.2	0.7	0.3	1	0

TABLE 6. Parameters of the chosen TAR(3) process.

N_{ls}	M	T_e	L	N_{BINS}	σ_ε	σ_η	T_{MAX}
729000	100	1200	200	128	0.0	0.5196	1200

TABLE 7. Parameters used in TAR(3) process based experiment.

parameters of the process and the initial conditions used to generate the data are presented in table 6.

A summary of parameters used in the combined forecast experiment is given in table 7.

5.1. Choice of parameters for the Threshold Autoregressive Process: construction of the data set and climatology. The TAR(3) time series that comprise the leaning set is first generated using the parameters in table 6 and equation 44. The resulting time series is displayed in figure 25. The series exhibits oscillating behaviour, varying between approximately -70 and approximately 0, in a periodic manner.

The climatology was then constructed using the learning set and $N_{BINS} = 128$ values bins, where the bins range from the minimum to the maximum vales observed in the leaning set (as in all other previous experiments). The resulting climatology is displayed in figure 26.

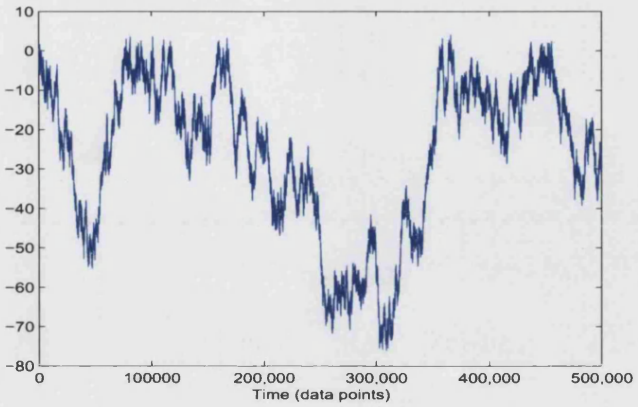


FIGURE 25. Threshold Autoregressive process order 3 (TAR3), x axis indicate time in generated data points and y axis is value.

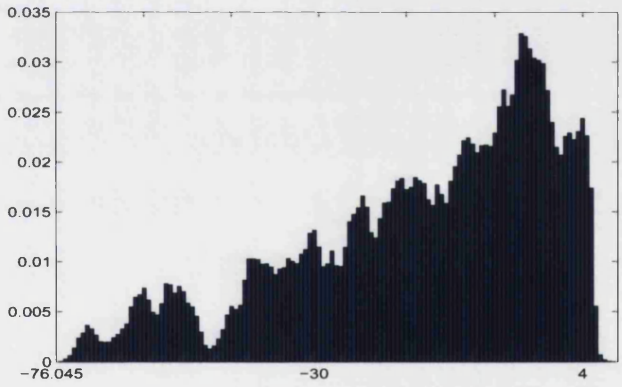


FIGURE 26. Numerically computed climatology of TAR3 constructed using 128 equally sized bins ranging between the max and the min values of the observed 729000 generated data points.

5.2. Initialisation of ensembles. Next $L = 200$ uniformly distributed locations on the length of the attractor (where $a = 1$, $b = 729000$, equation 27) are chosen and perturbed.

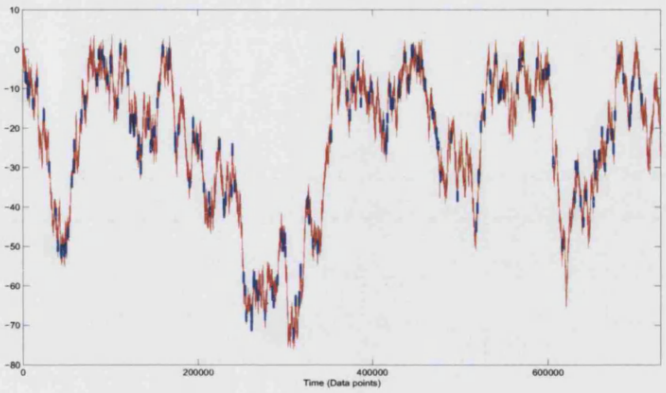


FIGURE 27. Perturbed initial conditions (blue) and their locations on the TAR3 time series (red) that has been generated using normally distributed noise with s.d. $= 0.519615$.

The observed behaviour of the ensembles is very different to those of the chaotic systems. In the TAR(3) time-series, the predictability of ensemble forecasts do not vary much in different locations, and each ensemble exhibits similar spread. The general behaviour of the ensemble is that the spread increases with time. Figure 28 displays three examples of ensemble forecasts obtained as a direct result of a perturbation of the initial conditions, in three different locations on the learning set.

As a result of ensembles being more steady in their behaviour, where general increase in the ensemble spread is observed, the values of α^t and σ_m^t should be much more smooth without jumps.

5.3. Parameter surface for the Threshold Autoregressive Process using ignorance skill score and it's minima. Indeed, when we consider the estimated parameters α^t and σ_m^t, figures 29 and 30 respectively, we can observe that the values of the parameters do not exhibit jumps. More

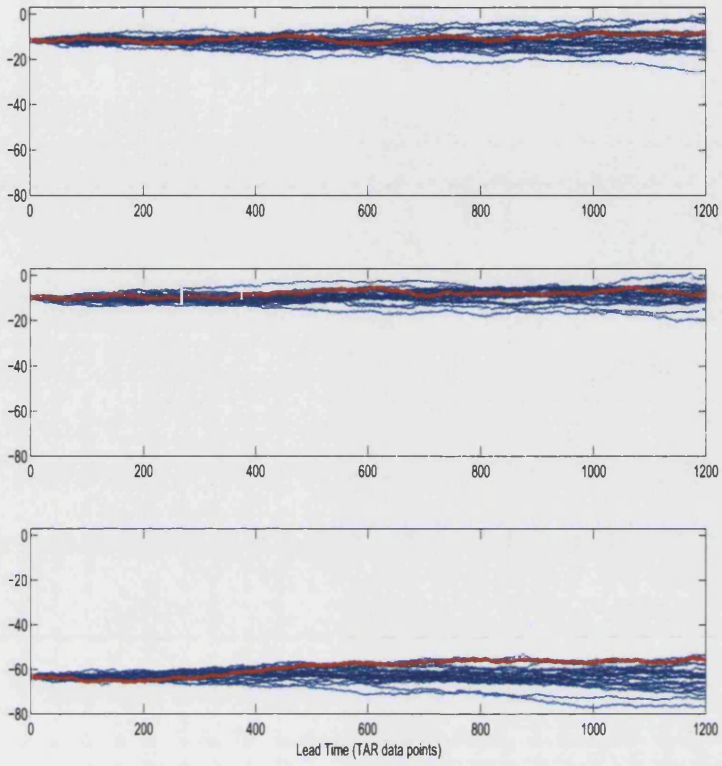


FIGURE 28. Examples of generated ensembles (blue), where different realisations of the random component of TAR3 has been used for each ensemble member, together with the verification (red) generated from perturbed initial conditions at given locations. x axis- lead time and y axis value.

than that, the values of α^t decrease at a relatively steady rate after 600 data points.

σ_m^t also increases steadily, but more steadily in the beginning, and then more steady at the level of 10, or approximately 13% of the spread between max and mean of the generated series. Interestingly, the observed values of σ_m^t exceed twice the value of the perturbation noise that has been used.

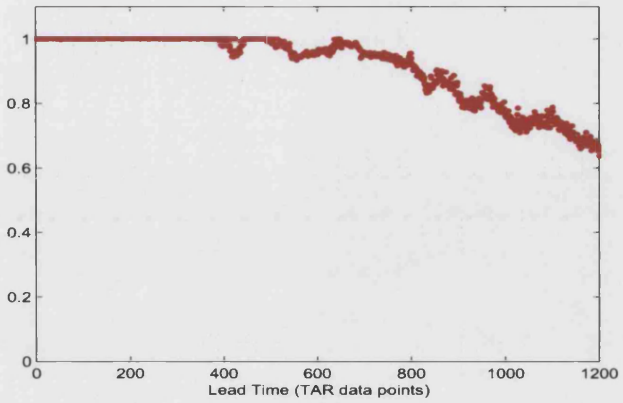


FIGURE 29. The blending parameter α^t obtained for TAR3.

This is consistent as the process contains more randomness and less structure than the chaotic systems, therefore larger levels of σ_m^t are expected.

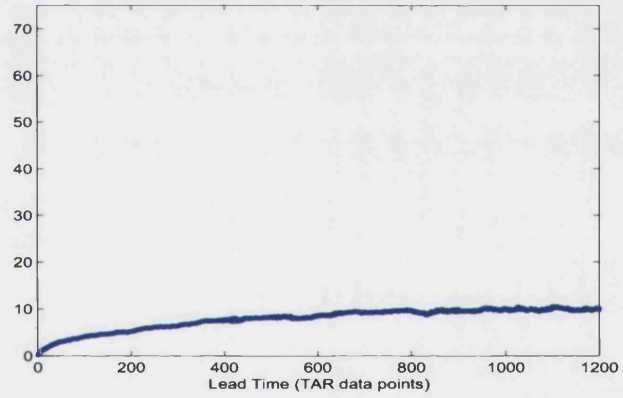


FIGURE 30. The kernel dressing parameter σ_m^t obtained for TAR3.

5.4. Analytical probability forecast verses the numerical approximated forecasts. Finally the combined probability forecasts of the TAR3 time series is produced, by perturbing an out of sample point of the series with the same level of perturbation noise as used in the estimation

of α^t and σ_m^t, i.e. $\sigma_\eta = 0.057735$. An ensemble forecast is then generated, and is shown in figure 31. It can be observed that the ensemble forecasts increase in spread steadily as the lead time increases.

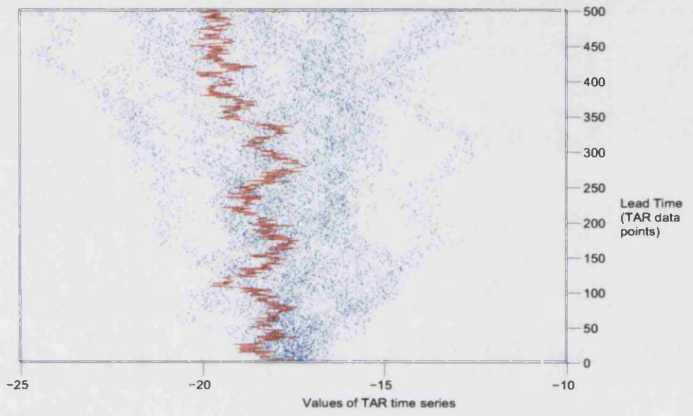


FIGURE 31. Ensemble forecast of TAR3 used in the construction of final probability forecast.

The resulting probability forecasts, computed using the estimated α^t and σ_m^t, together with the ensemble forecasts and the predetermined climatology, is shown on figure 32.

As expected the prediction given for TAR(3) is much more vague in its nature, compared to the ensemble prediction that was produced in the chaotic systems study. It is evident that even for a very short lead time the uncertainty in the forecast is huge.

Finally lets compare the numerically obtained prediction with the analytically computed one. We can calculate the analytical probability density function of the one, two and three day forward forecast. As the TAR3 process, defined in equation 44 contains a normally distributed random component, the forecasting pdf is also in the form of a normal distribution. The

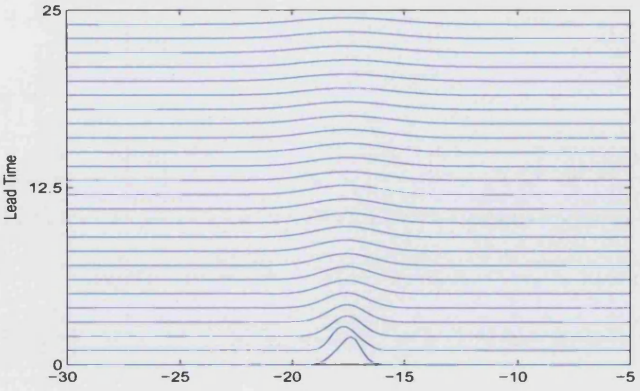


FIGURE 32. Final probability forecast issued for the time series of the TAR3 process, where x axis indicate the value of the series and y axis is lead time.

mean μ^d_{TAR3} and variance σ^d_{TAR3} of the forecasting pdf for one $d = 1$, two $d = 2$ and three $d = 3$ day forward forecast of the TAR3 time series are presented in equations 45, 46 and 47 respectively.

$$\begin{aligned} \mu^1_{tar3} &= C_1 + a_1 x_t + b_1 x_{t-1} + c_1 x_{t-2}; \\ \sigma^1_{tar3} &= \sigma_{\xi^1_t}; \end{aligned} \tag{45}$$

$$\begin{aligned} \mu^2_{tar3} &= C_1 (1 + a_1) + \left(a_1^2 + b_1\right) x_t + \\ &+ (a_1 b_1 + c_1) x_{t-1} + a1 c_1 x_{t-2}; \\ \sigma^2_{tar3} &= \sigma_{\xi^1_t} \left(1 + a_1^2\right); \end{aligned} \tag{46}$$

$$\mu^3_{TAR3} = C_1\left(1+a_1+a_1^2+b_1\right)+x_t\left(a_1^3+2a_1b_1+c_1\right)+$$

$$+ \left(b_1a_1^2+a_1c_1+b_1^2\right)x_{t-1}+\left(a^2c_1+b_1c_1\right)x_{t-2};$$

$$\sigma^3_{TAR3} = \sigma_{\xi_t^1}\left(1+a_1^2+\left(a_1^2+b_1\right)^2\right); \tag{47}$$

These forward probability forecasts are presented in figure 33, in red, together with the numerically obtained combined forecasts, in blue.

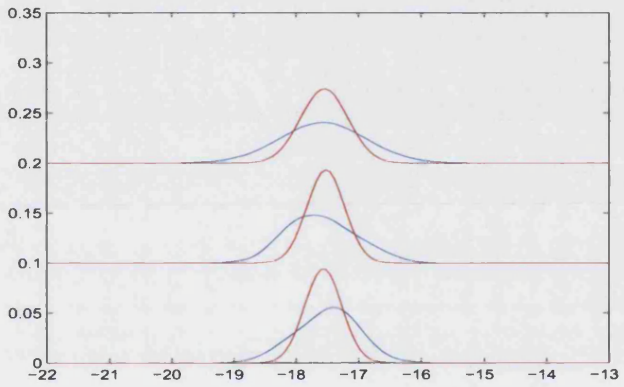


FIGURE 33. Comparison of the analytical one, two and three day forward probability forecasts (red) with the numerically obtained combinational forecasts (blue).

It can be seen that the numerical distributions are skewed compared to the analytical ones. As time increases the numerical pdfs become more normal in their shape. The skewness can be corrected by increasing the number of the ensemble members.

5.5. Conclusions of the experiment. It has been observed that in the case of chaotic systems, if the variance of the perturbation noise is smaller than the variance of the observational noise, the levels of σ^t remain

close to the size of the perturbation noise. When the verification is outside the ensemble α^t drops in value, but because the ensemble splits (when some of the members go to a different wing of the attractor), and hence some of the ensemble members still capture the move in the verification data, α^t does not drop significantly. When ensemble members recombine the value of α^t jumps back to one. Even after 7.0 Lorenz seconds, the model has a lot of skill.

As the size of the perturbation noise increases the value of α^t drops to lower levels, but the general behavioral patterns of alpha remain. It can also be seen that the return to one happens much less frequently. For larger perturbation noise σ^t is also on average larger. More than that, when the size of the perturbation is doubled, the level of σ^t on average doubles as well, and again grows with time. The level of σ^t is lower than the variance of the perturbation noise.

In terms of ensemble size, the smaller the ensemble, the smaller the average levels of α^t, at a given time. This is because α^t is sensitive to whether the verification points are contained within the ensemble spread (between the minimum and the maximum values of the ensemble members on a given day). Again, σ^t increases as the ensemble size reduces. Doubling the ensemble size reduces σ^t (on average) by half for a given lead time. Finally, when the Moore-Spiegel experiment was run for much longer times, it was observed that α^t becomes more volatile with time. σ^t remained very stable, and very small.

With the random process, the level of sigma is much larger, as expected , due to the nature of the system. But the overall behavior of σ^t is very stable. α^t is also rather stable and remains at one for much shorter time, compared to the chaotic examples.

As mentioned previously, the optimisation of σ^t and α^t was done simultaneously. It would be interesting to investigate how the behavior of both change when this optimisation is performed independently.

CHAPTER 5

ERAP: chaos, synthetic weather and real weather data.

1. Overview

In this chapter a new weather generator is presented. Based on Random Analog Prediction [**50**], Ensemble Random Analog Predictions (ERAP) generates a collection of synthetic time-series, consistent with both the historical distribution and the short range (non-linear) dynamics of observations.

One of the goals of this thesis is to apply synthetic weather data, produced by a weather generator and in particular Random Analog Prediction (ERAP), to price weather derivatives (chapter 6). In general, whether the interest lies in pricing weather derivatives in different areas of finance and risk management, or perhaps in different application areas such as logistics, health, tourism, transport etc, estimates for the occurrence of various types of weather over a time period are required. Traditional statistical methods which use linear models are typically ineffective in estimating this information from limited observations. The ERAP approach constructs synthetic data for both non-linear processes as well as linear processes. If the model class is known to be linear (that is, if the unknown process that generated the data is known to be a linear process), well established linear techniques would probably outperform the ERAP approach. Both data analysis and

physical intuition suggest that the weather process of interest is not linear (chapter 2).

In order to assess the performance of the ERAP approach, in the second part of this chapter the statistics of the generated synthetic data are compared to the statistics of the input data. Testing the ERAP method on data sets of varying lengths and varying characteristics, generated by known non-linear processes with well understood characteristics, allows us to conduct a controlled assessment of the method. Two non-linear processes have been chosen to test ERAP, including: the weather like non-linear process generated in chapter 3 and the Lorenz data (see chapter 4) with added seasonality in a form of a sine wave. Finally ERAP is applied to the real weather data set, described in chapter 3, to produce synthetic weather data.

The terminology used in this chapter is defined in table 1.

2. The ERAP mechanism.

The ERAP approach can be though of as a dynamic re-sampling technique. In particular, the method is structured by finding similar patterns to the present state in the observed data. ERAP works on multiple time scales, in particular it identifies similar patterns both the shorter and longer time scales. The idea behind this is that once the long term similar patterns are identified, short term similar patterns are chosen from those long term ones. This multi-scale pattern search does not have to be limited to two time scales. Depending on the data set, a multi-scale approach could be adopted.

ERAP - Ensemble Random Analog Prediction	Ensemble Random Analog Prediction is the approach based upon the Random Analog Prediction (or RAP, see L.A. Smith 1997) that has been extended to the ensemble mode.
synthetic data	Data produced by ERAP method.
system	System is an assemblage of inter-related elements comprising a unified whole, and in this document is described by a mathematical model.
robustness	Robustness in this document is referred to, in some sense, consistency of the statistical estimations.
perfect model scenario	Controlled experiment, where a data used has been generated from a chosen system with known characteristics.
learning set	A learning set is data that serves as an input into the ERAP method.
'true' statistic	statistic of the system
true PDF	An invariant measure of the system

In this section the step by step algorithm of the ERAP technique is described. Parameter definitions are given in table 2.

switch of the Lorenz system	The switch occurs when points switch from one wing of the Lorenz attractor to another.
number of consecutive decreasing / increasing days	Each set of the consecutive increasing or decreasing days is considered as a unique set, i.e. the number of days in a set is counted only once. For example, if a set contains 5 points, the number of 5 consecutive days is recorded once and the number of 4,3,2 consecutive days are not calculated from this set.
Value At Risk (VAR)	Value at Risk is defined to be a value V, such that all the observed values only go above(below) that point a $p\%$ of the time.

TABLE 1. Glossary of terminology and definitions.

Consider a given time series $\{s_i\}$ where $i = 1 \ldots N$. Before the method is applied to the chosen time series, a size for the *Learning Set* N_l is chosen, where *Learning Set* is defined to be the set of data which will be used as an input into the model.

(1) First stage of the method is to arrange the time-series into a time delay space [**27**] of a chosen size, as shown in equation 48.

s_i	An observation in a chosen time series.
N	The number of points in a chosen time series.
N_l	Length of the learning set.
T_0	The last observation in the learning set.
m_{long}	Space delay parameter:size of the window length for the long term scale.
m_{short}	Size of the window length for the short term scale.
τ_{long}	Time delay parameter- long time scale.
τ_{short}	Time delay parameter- short time scale.
ζ_1	A number that defines first constraint of the *Relevant set.*
X	The delay matrix of the chosen time series.
$m_{reduced\_long}$	Number of components in the reduced space.
$m_{reduced\_short}$	Number of components in the reduced space.
$N^{long}_{Neighbours}$	Number of nearest neighbours for the long term scale.
$N^{short}_{Neighbours}$	Parameter of ERAP: number of nearest neighbours for the short term scale.
$Max_{extention}$	The maximum number of points by which data can be extended at each iteration.
ϖ_j	Weight assigned to each short term neighbour.

(48)
$$X = \begin{pmatrix} s_1 & s_{\tau+1} & s_{2\tau+1} & \cdots & s_{m_{long}} \\ s_{\tau+1} & s_{2\tau+1} & \cdots & \cdots & s_{m_{long}+\tau} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \ddots & \cdots & \cdots & \cdots & s_{N_l} \end{pmatrix}$$

τ	Delay parameter.
$\alpha_{iteration}$	The number of points of synthetic data constructed at each iteration.
$Min_{extention}$	Minimum number of days by which data is extended on each iteration
$Max_{extention}$	Minimum number of days by which data is extended on each iteration
λ_i	A singular value
C	The transformation matrix of SVD

TABLE 2. The definitions of the parameters of the ERAP algorithm.

where τ is the delay parameter and m_{long} is size of the long time-scale window (see table 2 for parameter definitions).

(2) Let the i^{th} row of the X matrix be denoted as x_i, such that:

$$x_i = \{s_{(i-1)\tau+1}, s_{i\tau+1}, s_{(i+1)\tau+1}, \cdots, s_{(i-1)\tau+m_{long}}\} \tag{49}$$

The technique works by finding similar patterns among it the x_i to the current state x^*. On the first iteration x^* will be the last m_{long} data points of the learning set and T_0 will be the last component of the x^* ($T_0 = s_N$). At every iteration of the method x^* is shifted such that newly generated data points are included in the new x^*.

(3) In order to make the search for the closest patterns more targeted, the rows of X are filtered, which means that every point in the constructed $\mathbb{R}^{m_{long}}$ space is filtered, and only the points that satisfy

a set of criteria will form the *'Relevant Set'* from which the closest patterns are chosen. The first criterion is that x_i belongs to the *'Relevant Set'* only if its last component is within plus or minus ζ_1 of T_0. Such points are identified at this stage and their indices are recorded.

(4) The next part of the method attempts to reduce the dimentions of X using to remove the predominantly noise influenced components. With SVD any matrix X can be represented as shown in the equation 50.

$$X = U\Sigma V^T \tag{50}$$

Where: U is a matrix containing *Left Singular Vectors*; Σ is a diagonal matrix containing *singular values* and V is a matrix containing *Right Singular Vectors.* The Singular vectors contained in U are found by taking the eigenvectors of XX^T, and the singular vectors contained in V are found by computing the eigenvectors of X^TX. V plays the role of a transformation matrix, which allows us to represent the matrix X in an orthogonal representation, where basis vectors are given by Singular Vectors. The singular vectors represent directions of maximum variance stretch; and singular values the amount by which that particular direction is stretched. SVD can be applied to any matrix. Furthermore, a 'cheap' SVD can be achieved by applying the decomposition to X^TX. As we only need V^T for further work we can show that applying SVD to X^TX

can produce desired V^T. In particular, given the decomposition of X defined in equation 50, then:

$$\begin{aligned} X^T X &= \left(U\Sigma V^T\right)^T \left(U\Sigma V^T\right); \\ &= V\Sigma^T U^T U\Sigma V^T; \\ &= V\Sigma^2 V^T; \end{aligned} \tag{51}$$

(5) Taking the dot product of X with only the first $m_{reduced\_long}$ number of rows of V, projects X into the lower dimensional space $R^{m_{reduced\_long}}_{Orthogonal}$, allows us to reduce the dimensionality and as a result the noise dominated components.

(6) In order to avoid running out of data points while constructing future trajectories, a further restriction is applied to exclude certain rows of X from the *'Relevant Set'* reducing it even further. The rows excluded are: $x_{N_l - m_{long}+1-Max_{extention}} \cdots x_{N_l - m_{long}+1}$, where $Max_{extention}$ is the maximum number of points that can be generated at each iteration.

(7) The next step is to find $N^{long}_{Neighbours}$ 'closest' points in $R^{m_{reduced}}_{Orthogonal}$ to x^* from the set of x_i. These nearest neighbours' are chosen from the *Relevant Set* only (the *Relevant Set* defined in 3rd and 6th steps of the method). Euclidian distance is used to define the closeness.

(8) Once the closest long time scale neighbours are found their indices are recorded. The *'Relevant Set'* used in the next steps, only contains those points which are the closest neighbours to x^* in the long-term analysis. For example, further analysis will only be conducted within the correct season.

(9) Many time series, a specially weather data exhibit long and short term behavioural patterns. In order to capture short term behaviour, once the nearest neighbours have been found, and the information concerning their location in $R^{m_{reduced}}_{Orthogonal}$ has been recorded, nearest neighbours for the short time period are also found. This is achieved by first constructing the m_{short} dimensional time delay space from the Learning Set, as given in equation 52.

(52)
$$X_{Short} = \begin{pmatrix} s_1 & s_{\tau+1} & s_{2\tau+1} & \cdots & s_{m_{short}} \\ s_{\tau+1} & s_{2\tau+1} & \cdots & \cdots & s_{m_{short}+\tau} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \ddots & \cdots & \cdots & \cdots & s_{N_l} \end{pmatrix}$$

x^* now will contain m_{short} number of elements.

(10) Singular Value Decomposition is again applied, in order to reduce dimension.

(11) X_{Short} is then transformed into the $R^{m^{reduced}_{short}}_{Orthogonal}$ space defined by the short term singular vectors. Again this is achieved by taking a dot product of the new X_{short} with the first $m^{short}_{reduced}$ components of new V.

(12) In the $R^{m^{reduced}_{short}}_{Orthogonal}$ space we find the $N^{short}_{Neighbours}$ closest neighbours, as before, by calculating, Euclidian distances between x^*_i and the rest of the points in the adjusted *'Relevant Set'*.

(13) Once the set of short-time nearest neighbours have been chosen, one nearest neighbour is picked at random with an assigned probability, given by equation 53, that is inverse proportional to the distance from the x^*.

$$\sum(\varpi_j/distance_j) = 1 \tag{53}$$

Where ϖ_j is the weight of a particular short term nearest neighbour j.

(14) We select $\alpha_{iteration}$, which represents by how many points data will be extended on this particular iteration, at random from the predetermined range, $\Re = [Min_{extention} \ldots Max_{extention}]$, where each extension length is an integer, and is equally likely.

(15) Then, the future trajectory of x^* is constructed using the future trajectory of the nearest neighbour, as in the equation 54.

$$T_{j+1} = T^{Neighbour}_{j+1} - T^{Neighbour}_{j} + T_j \tag{54}$$

where $j = 0 \ldots \alpha_{iteration}$. $T^{Neighbour}_0$ is the last component in the chosen closest neighbour vector and $T^{Neighbour}_j$ is the j^{th} component of the future trajectory of the nearest neighbour. In other words, the one day forward 'prediction' is defined by the sum of the

last known data point and the first forward difference of the future trajectory of the nearest neighbours.

(16) The above procedure is repeated to generate a time series that constructs one member of the synthetic data ensemble. Finally, we repeat the steps, starting with the same learning set and the same parameters, to generate different realisations, producing an ensemble of synthetic weather data.

2.1. Parameters in more depth. In this section some brief reasoning and intuition behind the initial choice of parameters is presented. Additionally a parameter calibration methodology is proposed, but this methodology was not implemented in this thesis. First let's consider what influence each parameter has on the generated ensemble.

- *The size of the learning set* N_l:

  The size of the learning set would depend upon the total amount of data available. A sufficient amount of data must be allocated for out of sample performance assessment. Additionally the size of the learning set must be large enough to capture long term patterns. As for any time series analysis techniques, ERAP heavily relies on the amount of data available in a learning set and the general rule remains, the more data available the better.

- *The size of the Reduced dimension for the long term* $m_{long\_reduced}$:

  This parameter allows us to reduce the 'numerical cost' of searching for nearest neighbours, by only searching using the least noise affected components. This parameter was chosen by adding noise

to the original time series and then arranging it into the same size delay space as the original time series. Then Singular Value Decomposition is performed. Singular values which are least affected by the added noise give an indication to the extent that the size of the dimension can be reduced.

- *The size of the Reduced dimension for the short term* $m_{short\_reduced}$:

  As for the long term case Singular Value Decomposition is performed on both the noisy learning set and the original after they have been arranged into the delay space using m_{short} . Singular values which are least affected by the added noise determine the size of the reduced dimension.

- *Number of long term neighbours* $N^{long}_{Neighbours}$:

  This parameter defines the size of the set, that contains neighbours exhibiting similar long-term behaviour to the present state. The size of the set has to be large enough to permit 'freedom of choice' but on the other hand it will serve as a filter, so further analysis will be based only on relevant information. Generally it will depend on the size of the data set, the magnitude of the delay parameter and the clustering of points in delay space.

- *Number of the short-term neighbours* $N^{short}_{Neighbours}$:

  In order to determine the size of the $N^{short}_{Neighbours}$ parameter, a Talagrand Diagram [**55**] has been used. In particular, procedure of ERAP has been slightly altered by proceeding through the method until step 13. At this stage instead of choosing the size of

the extension of the time series at random from the chosen range: $Min_{extention} \ldots Max_{extention}$, a predetermined fixed size extension will be used on every iteration, for every ensemble member. This results in an equal number of one-day, two-day, three-day, etc forward ensembles in the constructed synthetic data. This construction of the synthetic data allows us to assess the quality of the one-day forward, two-day forward etc, frequency distribution, produced by ensembles with a chosen $N^{short}_{Neighbours}$ parameter. The size of the fixed extension used for this exercise is equal to 9 days.

After completing the construction of the synthetic data with a fixed extension size, a Talagrand Diagram can be assembled. First only the one-day forward ensemble of realisations is considered. The realisations are arranged into increasing order. Then boundaries of the Talagrand bins which are determined by the predictions themselves, are established. In particular: Bin1: $(-\infty; T_1)$; Bin 2: $[T_1; T_2)$; Bin 3: $[T_2; T_3), \cdots$, Bin 33: $[T_{32}; \infty)$ on a given simulation day. T_1 is the smallest (for the weather data coldest) realisation T_2 is the second smallest, and so on. This allocation of the bin size is produced for each one-day forward ensemble. Then for each one day ahead forecast the actual temperature is recorded in relation to the pre-defined bins. In particular, we see which bin the actual observation falls. This is conducted for all one day ahead forecasts. So we get a frequency diagram. This is repeated for all days ahead. So for each day ahead we a Talagrand Diagram.

Ideally, the Talagrand diagram will be flat, which will indicate that the distribution of the real data was sampled well with a chosen number of scenarios in the ensemble. The bins are chosen to ensure that there ought to be an equal probability of a measurement falling into each bin. If the diagram is not flat, then the probability distribution implied by the ensemble is not well represented.

Once the Talagrand diagram has been constructed forecast day ahead the deviations from the mean bin level are computed by deriving the Υ as shown in the equation 55. When $\Upsilon = 0$ a histogram is flat:

$$\Upsilon = \sum_i (b_i - \mu^T)^2 \tag{55}$$

Where b_i is the number of observations in i^{th} bin and μ^T is the mean bin level, which is defined as the average number of points in a bin: $\frac{Total_Number_Obs}{N_bins}$.

If the number of points in the outside bins is considerably large, $N^{short}_{Neighbours}$ could be increased, as one of the reasons for overcrowding of the outside bins might be the small size of the ensemble, resulting in missing a large proportion of the tails of the real data's distribution. If, on the other hand, too many points are concentrated in the middle bins of a histogram $N^{short}_{Neighbours}$ could be reduced, as one of the reasons could be that the range of values produced by the ensemble is too large.

- *First restriction in construction of the 'Relevant Set'* ζ_1:

This is the first restriction which is applied to the Learning Set, after it has been arranged into the time delay space. In order to make the search more targeted, only rows x_i of the matrix X the last component of which is within plus or minus ζ_1 degrees of T_0 will be considered as candidates for the closest neighbours. The value of this parameter can be determined by considering the difference between the observed maximum and minimum values in the learning set, and choosing ζ_1 such that the search for the nearest neighbours is conducted only within, say, 10% (with respect to the size of the of the difference between max and min values observed in the learning set), of T_0.

There could be other restrictions that allow us to construct a more targeted search for the nearest neighbours, such as phase control of the nearest neighbours. For example, the neighbours can be forced to be chosen from similar times of the year, However, it is important to allow the method to have enough freedom to choose nearest neighbours, especially for the long term phase of the procedure, from 'out of character' locations. For example, a given April can be more like February or sometimes like May. The phase control restriction must allow such choice.

- Finally we discuss methods that can be used to determine: Long-term Lag - m_{long}, Short-term lag - m_{short}. The maximum and minimum number of days in the forward trajectory - $Min_{extention}$ and $Max_{extention}$. The choice of m_{long} is driven by the desire to

capture the most dominant long term pattern present in time series, such as seasonality (for weather data). Parameter m_{short} captures short-term behaviour of the system. As for the long term lag, the initial choice is currently based on intuition and nature of the time series. $Min_{extention}$ and $Max_{extention}$ are parameters that determine the size of the chunks of the historical data that make up generated synthetic data at each iteration. Choosing the value of the $Min_{extention}$ parameter too small would prevent capturing of the behaviour of the real weather variables trajectory, and choosing the value of the $Max_{extention}$ parameter too large will result in a poor quality of the synthetic weather data. These four parameters could be optimised by minimising Ignorance Skill Score [**45**] (described in chapter 4), where all the other parameter of the ERAP method are fixed. This optimisation is not conducted in this thesis.

3. Testing ERAP: Experimental Design in the perfect model scenario.

This section describes the experiment that has been carried out on synthetic data generated by the ERAP method. The main goal of the experiment is to strengthen the argument that the data generated by ERAP provides better estimates for the 'true statistics of the underlying system, compared to the statistics calculated from a limited data set of that system. Additionally, it can be determined how well ERAP can quantify the uncertainty in estimates of statistics compared to limited observations.

The experiment is conducted on two constructed data sets: Lorenz (chapter 4) and 'weather-like' data (chapter 3). In this case, Lorenz data is the output of the Lorenz system's ([**44**]) x component that has been superimposed onto a sine wave. The sine wave has been added so that yearly periodicity and seasonality characteristics of temperature data are present to some extent [1]. Finally the ERAP approach is applied to the Berlin data described in chapter 3, to produce an ensemble that is later used in chapter 6 to price a weather derivative.

The experiment is constructed as follows:

(1) Generate experimental data that includes the actual (sometimes refereed to as the 'truth' or verification) and also the learning set. The number of points generated is referred to N_{actual}.

(2) Determine the parameters of ERAP in order to generate synthetic data using the chosen learning set, where the number of points in the learning set is defined as N_l.

(3) Generate ERAP ensembles using chosen learning sets. Different sizes of learning set are used in the case of 'weather-like' process.

(4) Calculate and compare traditional statistics, including frequency distributions and distributions of moments of the ERAP ensemble to actual frequency distribution and moment (frequency distribution and moments computed using the entire data set with N_{actual}

[1]Seasonality not only manifests itself in the rise and fall in the summers and winter. There are other characteristics such as seasonal daily fluctuations etc., that will not be represented by just a sine wave.

points) and also the frequency distribution and moments of the learning set (computed using N_l data points).

(5) Calculate and compare the distribution for the number of consecutive 2,3,4...6 cold/warm days of the ERAP ensemble to actual and learning sets based distributions of consecutive days, (consecutive cold/warm days are defined as monotonically decreasing/increasing days in the data).

(6) Calculate and compare distributions of number of freezing days for all data sets.

(7) For the Lorenz data, calculate and compare the proportion of switches (refer to table 1 for the definition of switches) for the learning set and the actual data, and the distributions of switches for the ERAP ensemble data.

(8) Compute and compare relative frequencies of lower and higher extreme values again based on the ERAP ensemble, the learning set and the actual data.

(9) Calculate and compare the percentiles for all data sets.

4. Controlled experiment: weather-like data.

8,000 years of weather-like data was generated as described in chapter 4. The statistic of these 8,000 years is considered to be the 'truth'. We consider two sizes of learning set: 1,000 and 100 generated years, where each year consists of 365 data points. These two learning sets are used to generate ERAP ensembles that are then compared, in terms of statistic calculated

from the ensembles, to the 'truth' and also to the statistics computed from the learning sets.

4.1. ERAP parameters for the weather-like process. As mentioned previously the parameters of ERAP that have been used in all the controlled experiments are initial choices that could be further optimised, as proposed in section 2.1. In this chapter we present an illustration of some of the concepts described in section 2.1, in order to demonstrate the reasoning behind the initial parameter choice. All the presented figures in this section were produced using the 1000 years learning set, the results for the 100 years learning set were very similar.

The weather-like data has many characteristics of real weather data (by construction). Hence m_{long} (size of the long term window) and m_{short} (size of the short term window) might reflect season and a particular week for example. That is to say, that one wishes to determine a similar season in the past data and then from that choice of similar season, one wants to pick the most similar week to the weeks immediately prior to the point of forecast.

First, the delay matrix X is constructed from the learning set and then SVD is performed. Singular values for both long and short term are given in figure 1 and 2 respectively.

Similarly, long and short term singular vectors are shown in figures 3 and 4 respectively.

In order to determine the nearest neighbours, the Talagrand diagram has been used as proposed in section 2.1. Here we present one day ahead,

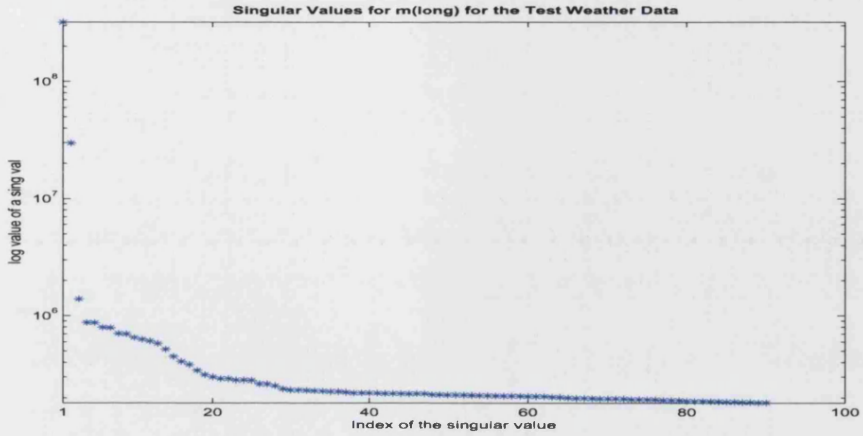


FIGURE 1. Singular values derived for the long term case constructed using 1000 years of the learning set of the weather-like process.

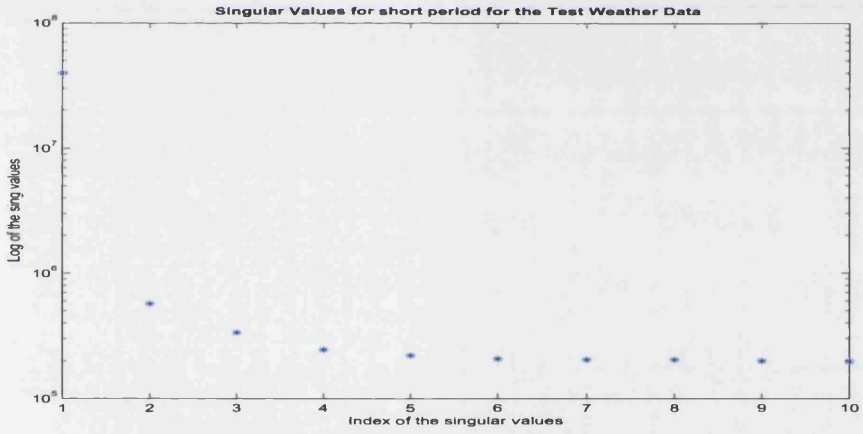


FIGURE 2. Singular values derived for the short term case constructed using 1000 years of the learning set of the weather-like process.

five day ahead and nine day ahead simulation based Talagrand diagrams in figures 5, 6 and 7 respectively.

Lets summarise the parameter values that were chosen for the ERAP generator applied to the weather-like process:

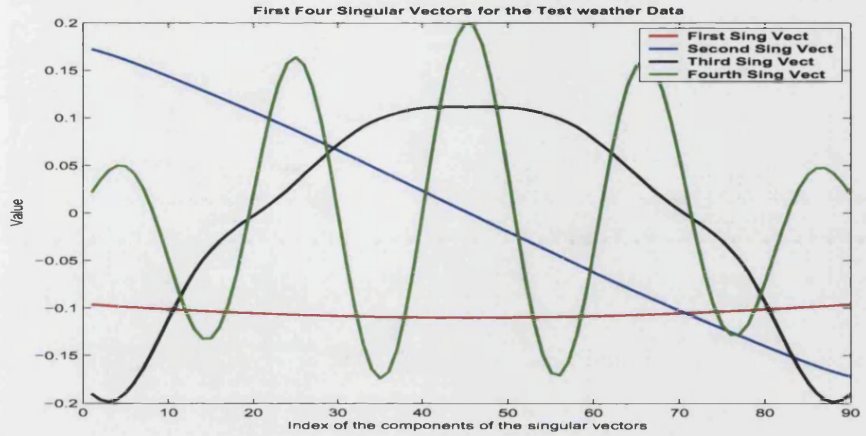


FIGURE 3. First Four Singular Vectors of the Weather-like Data, derived for the long term case.

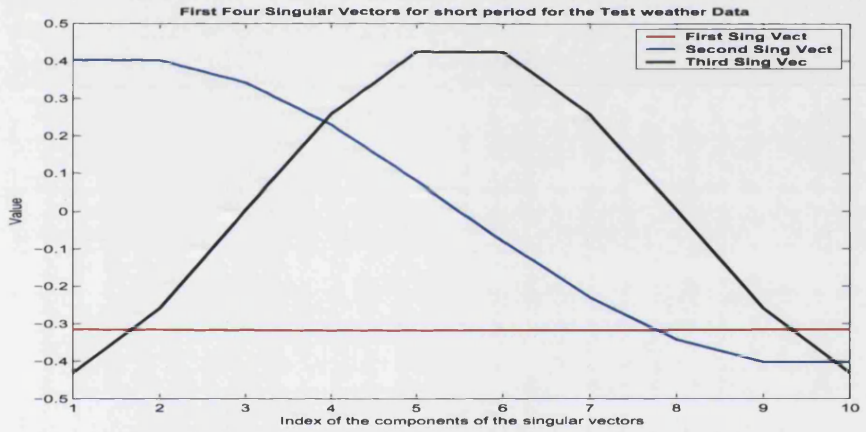


FIGURE 4. First Four Singular Vectors of the Test Weather Data, derived for the short term case.

Note that in this case we used both: the value restricting parameter ζ_1 and the phase control parameter that allowed us to conduct a more targeted search. This results in a stable ensemble of synthetic data.

4.2. ERAP ensemble for the weather-like process. Once all the parameters had been chosen, the ERAP weather generator was run. The resulting synthetic test weather-like data is presented in figure 8. In particular,

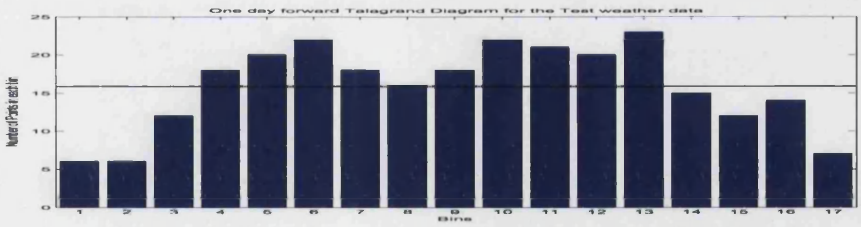


FIGURE 5. Talagrand Diagram for the One Day forward simulations for the Test weather data. Horizontal magenta line indicates the sampled mean of the histogram

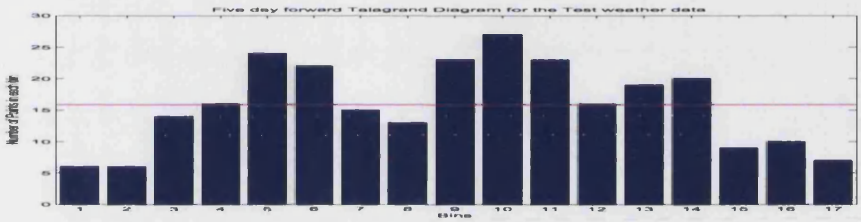


FIGURE 6. Talagrand Diagram for the Five Day forward simulations for the Test weather data. Horizontal magenta line indicates the sampled mean of the histogram

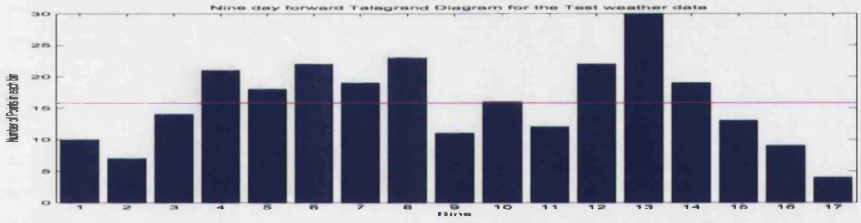


FIGURE 7. Talagrand Diagram for the Nine Day forward simulations for the Test weather data. Horizontal magenta line indicates the sampled mean of the histogram

the first two rows (blue) illustrate individual ensemble members generated by the ERAP approach using 1,000 years as a learning set. Two samples of the verification data are illustrated in the third row in red.

N_l	36500 and 365000
τ	1
m_{long}	90
m_{short}	10
$m_{reduced\_long}$	4
$m_{reduced\_short}$	3
ζ_1	2
$No^{long}_{Neighbours}$	128
$No^{short}_{Neighbours}$	32
$Max_{extention}$	12
$Min_{extention}$	5
$\alpha_{iterations}$	100
$Size\_of\_ensemble$	100
$seed$	69
$phase\_control\_parameter$	5

TABLE 3. Parameters of ERAP for the Weather-like process experiment.

The whole ERAP ensemble for the weather-like process is shown in figure 9. The ensemble generated by the ERAP approach (blue) using 1000 years of the weather-like data as a learning set, the corresponding verification is shown in red.

4.3. Comparing statistic: weather-like process experiment. In the case of the controlled experiments, one knows the true statistic. In particular for the weather-like system, the true statistic is approximated by

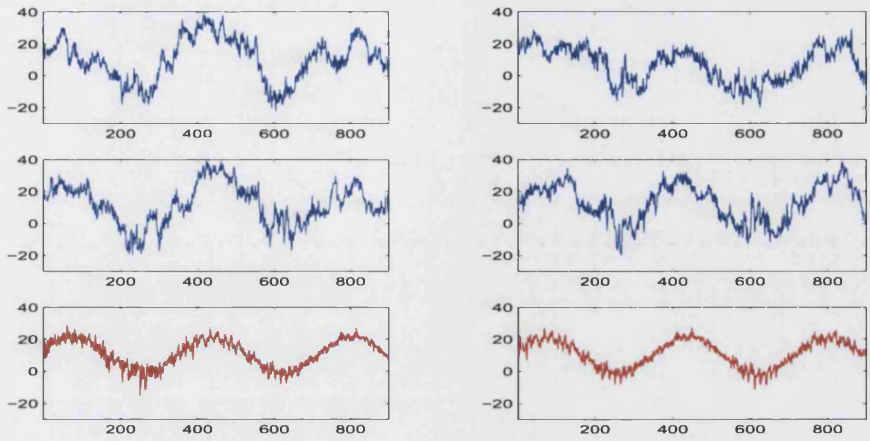


FIGURE 8. First two rows (blue) illustrate individual ensemble members generated by the implemented ERAP approach using 1,000 years as a learning set, two samples of the verification data are illustrated in the third row in red.

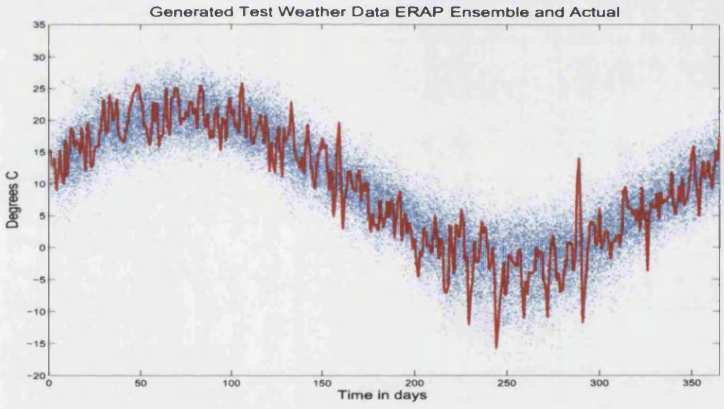


FIGURE 9. Ensemble generated by the ERAP approach (blue) for the learning set consisting of 1,000 years of the weather-like data. The verification is shown in red.

the 8000 years of weather-like data. 8000 years of the generated weather-like data will contain all the patterns that the process contains (please refer to section 3 for the details of the weather-like process). Each statistical

measurement in this section is therefore computed for the 8,000 years of the weather-like data (referred to as the 'truth', verification or the actual); two learning sets (1,000 years and 100 years of data); and the ERAP ensembles that resulted from these two learning sets. The ERAP based statistic is always presented in the form of a frequency histogram, where the bins are populated by the statistical measurements computed from each ensemble member.

We follow the steps of the experimental design outlined in section 3 in order to identify whether ERAP generated synthetic data is consistent with the learning set, and how well it represents the true statistics of the chosen system.

First we consider the 1st four moments [**48**]. The first moment is presented in figure 10. The mean of each ensemble was computed and then arranged into a relative frequency histogram which is given in blue on the plot. The mean of the learning set, which in this case contained 1000 years of data, is computed and presented in green. Finally we also compute the 'true' mean.

It can be seen that the first moment of the learning set and the actual first moment (the 'true' mean) are both well represented, and both the learning set mean and the actual mean fall into the maximum probability bin. The distribution of the mean based on ERAP synthetic data appear to be skewed. We also observe that the leaning set mean is identical to the actual mean, which tells us that the learning set is large enough to capture precisely the first moment statistic.

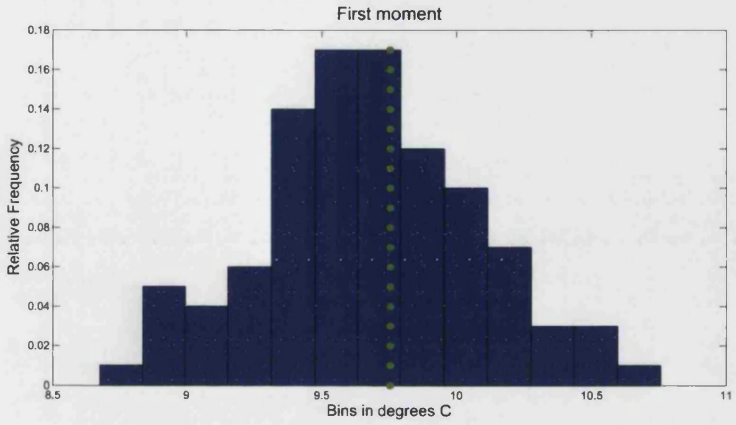


FIGURE 10. Comparison of the distribution of the first moment of the ERAP ensemble (blue) to the first moment computed from the 1000 years of the learning set (green) and the actual data based mean (red).

Now lets consider the standard deviation statistic, figure 11, where again synthetic ERAP data produces a distribution for the standard deviation (presented in blue), which is then compared to the actual s.d (red) and the learning set based s.d (green). This is a very important statistic for hedging weather risk, and in particular weather derivative valuation (see chapter 6 for more details on how variance is used in the process of weather derivative pricing).

It is evident that the learning set and the actual standard deviation are almost identical, and so again the learning set of 1000 years represented the 'true' standard deviation well. As a result, we expect the ERAP based standard deviation to be consistent with the true standard deviation. It is evident that this is the case, as the learning set standard deviation and

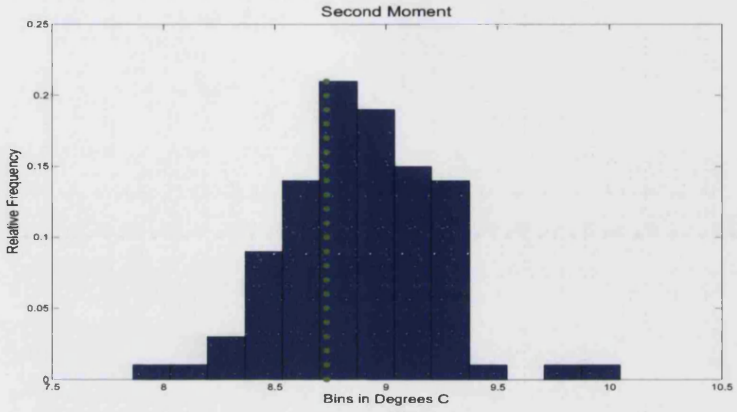


FIGURE 11. Comparison of the distribution of the standard deviation of the ERAP ensemble (blue) to the s.d computed from the 1000 years of the learning set (green) and the actual data based (red).

the actual standard deviation fall into the maximum probability bin of the synthetic standard deviation distribution.

Figure 12 illustrates the third moment of the synthetic ERAP data, presented in the form of relative frequency, together with the actual and the learning set third moments.

Both, the learning set and the actual skewness have small negative values, which is consistent with the synthetic skewness histogram which has most of its probability weighting allocated to the negative values. The actual and the learning set statistics are on the edge of the maximum probability bin. Again the learning set well represents the true skewness of the process.

Figure 13 demonstrates the synthetic distribution of the fourth moment (in blue), the fourth moment computed from the learning set that contains 1000 years of data (in green) and the actual 4th moment (red). As with the

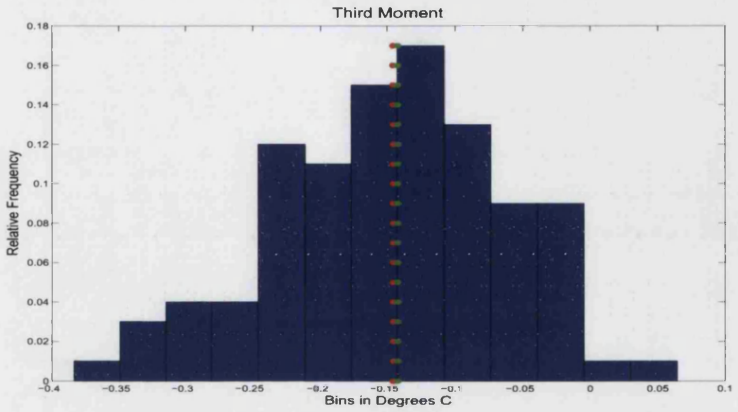


FIGURE 12. Comparison of the distribution of the third moment of the ERAP ensemble (blue) and the third moment computed from the 1000 years of the learning set (green) and the 'true' third moment (red).

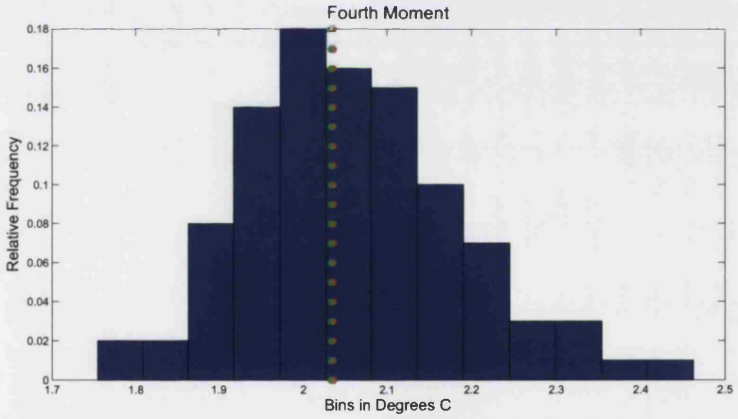


FIGURE 13. Comparison of the distribution of the fourth moment of the ERAP ensemble (blue) and the fourth moment computed from the 1000 years of the learning set (green) and the actual data based (red).

other moments, the learning set captures the actual kurtosis and hence the synthetic data kurtosis distribution is expected to be consistent with the

true kurtosis. We can see that although the true kurtosis does not fall into the maximum probability bin, it is well represented.

Now let's consider the yearly percentile. For the synthetic ERAP generated data, we compute daily 95th and 5th percent percentiles, presented in blue on figure 14. Then it is compared to the verification data (in red). We can see that the verification data goes outside the 5th and 95th isopleths approximately 10% of the time.

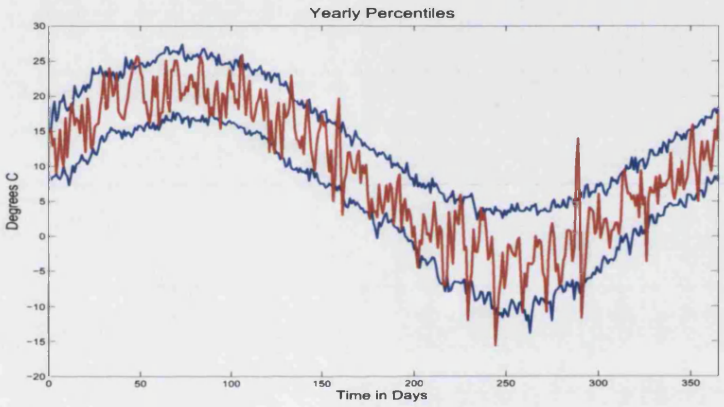


FIGURE 14. Comparison of the 95th and 5th isopleths' daily evolution generated by the ERAP ensemble (blue), and the actual data (red).

Figure 15 displays the 95th percentile relative frequency distribution computed from the synthetic ERAP data (blue) which is compared to the 95th percentile of the learning set (green) and the actual 95th isopleth (red). The relative frequency distribution is constructed by computing the 95th percentile of all the values observed in each ensemble member. Then all the ERAP based measurements are arranged into a relative frequency histogram.

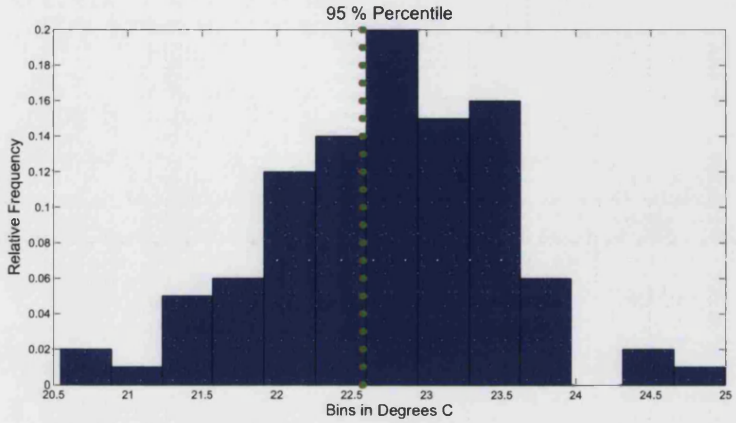


FIGURE 15. Comparison of the relative frequency of the ERAP based yearly percentiles (blue), 1000 years sized learning set based (green) and the actual yearly percentiles (red).

Similarly, the 5th percentile relative frequency histogram is constructed from synthetic data and compared to the actual and the learning set 5th percentiles. In both cases: the actual 95th and 5th percentiles are identical to the learning set and are well captured by the ERAP synthetic data.

Next we consider the consecutive increasing and decreasing days count, which was introduced in section 3. Here we consider the cases of 1,2,..,6 consecutive days. The ERAP synthetic data based measurements are presented in the form of a relative frequency histogram, after the consecutive increasing/decreasing days count has been conducted for each ensemble member. Such relative frequency histograms are presented in figures 17 for consecutive decreasing days and 18 for consecutive increasing days. The plots also illustrate the actual consecutive increasing and decreasing day count (red), and the count based on the learning set (green), adjusted to take into the account the number of years in the learning set and the actual. So one

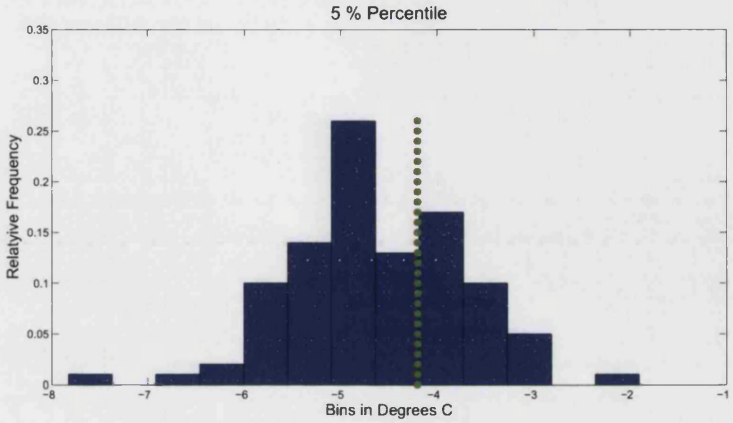


FIGURE 16. Comparison of the relative frequency of the ERAP based 5th % yearly percentile (blue), 1000 years sized learning set based (green) and the actual yearly percentiles (red).

might think of this statistic as an average per year count. Additionally we also present the mean of the synthetic data based count, which is illustrated in yellow.

One can see that in the case of 1,2,..., 5 consecutive cold days the learning set based measurement coincide with the actual count. In this case the actual count falls in the area of maximum probability on the synthetic data based relative frequencies. For 5 consecutive decreasing days, the actual count falls in an area where there is no probability weight, however the neighbouring bins exhibit the largest relative frequency. This is due to the large number of bins chosen to represent that frequency distribution. As number of consecutive days is increased, the number of observed occurrences decreases, and hence the number of bins in the construction of such a histogram should be reduced.

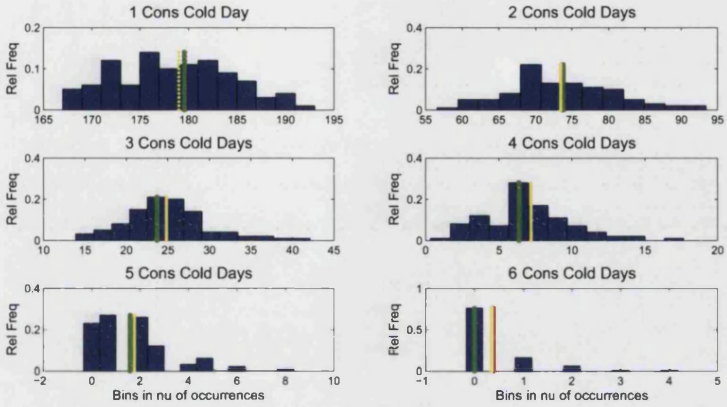


FIGURE 17. Comparison of the relative frequency of the ERAP based consecutive decreasing days (blue), together with the mean on that distribution (yellow), 1000 years sized learning set based (green) and the actual consecutive decreasing days count (red).

As the number of consecutive decreasing days being considered increases the statistic becomes harder to estimate, due to the rareity of such occurrences in all data sets. It can be seen that for 6 consecutive cold days, figure 17, the mean of the ERAP based distribution (yellow) coincides with the actual measurement (red), and the learning set based count (green) is lower than the actual and falls into the maximum probability bin. A similar result is observed for the consecutive increasing (or 'warm') days count, figure 18.

This is another statistical measurement that is important to be able to replicate well, especially as the number of consecutive days count increases. It is particularly significant, that this measurement is hard to estimate from limited historical data. It is encouraging that the ERAP based frequency was consistent with the learning set measurement (the learning set count fell

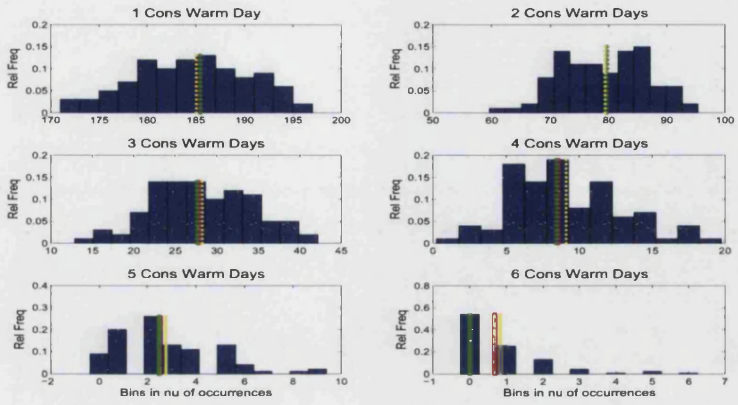


FIGURE 18. Comparison of the relative frequency of the ERAP based consecutive increasing days (blue) together with the mean on that distribution (yellow), 1000 years sized learning set based (green) and the actual consecutive decreasing days count (red).

into the maximum probability bin) and that the mean of the ERAP based relative frequency distribution coincided with the actual count. This tells us that the ERAP synthetic data produces enhanced statistics that are not necessarily captured by the learning set.

Next we consider the number of freezing days count, where freezing days are defined as values below or equal to zero. As with the consecutive days count, the number of freezing days is counted for each ensemble member. Then these measurements are arranged into a relative frequency histogram, see figure 19 in blue. Finally the actual freezing days count and the learning set based count are compared to the synthetic data based frequency and to each other. Again, the learning set count well represents the true freezing days count, both of which fall close to the maximum probability bin. It can

be seen however, that the ERAP based freezing days count overestimates the number of freezing days. This could probably be improved if the parameters of the ERAP approach are optimised, as mentioned in section 2.1.

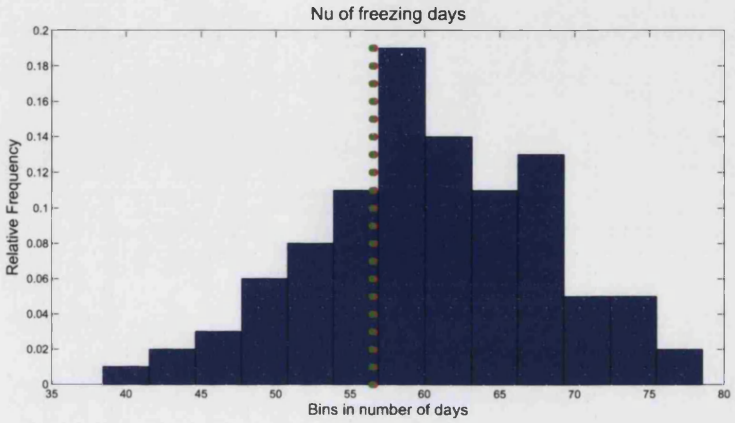


FIGURE 19. Comparison of the relative frequency of the ERAP based freezing days count (blue) to the freezing days count based on the 1000 years sized learning set (green) and the 'true' freezing days count (red).

Finally we consider the distributions of extreme low and high values. In other words we pay particular attention to the tails of the relative frequency distribution. Figures 20 and 21 show the distributions of extreme low and high values of the synthetic ensemble data (in blue where all ensemble members make one frequency distribution), the learning set (in green) and the actual (in red).

Once again, the learning set in this case well represents the measurements of the actual data set. As a result we expect the synthetic data based relative frequency distribution to represent the 'true' distributions well. It can be seen that the low and high value distributions are represented well,

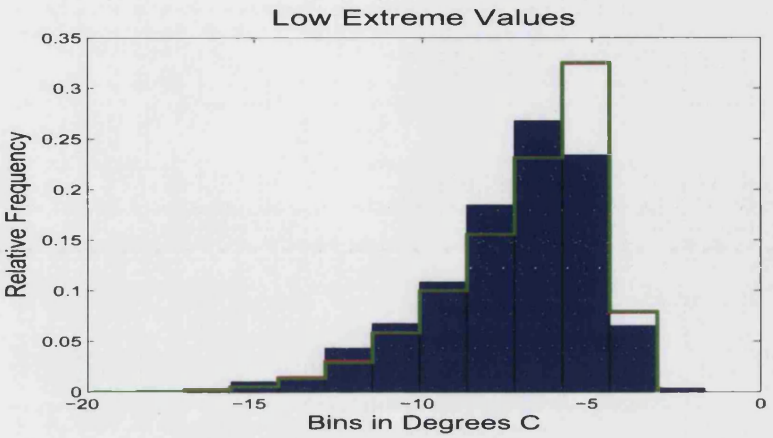


FIGURE 20. Comparison of the of the ERAP based (blue), relative frequency of the low extreme values computed from 1000 years based learning set (green) and the actual data based (red) relative frequency distributions of extreme low values.

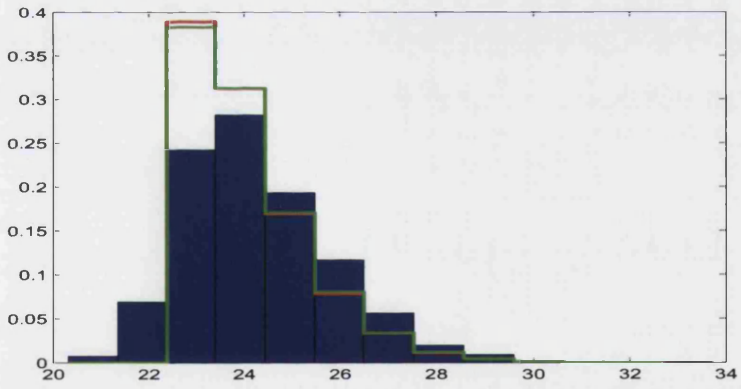


FIGURE 21. Comparison of the of the ERAP based (blue), relative frequency of the extreme high values computed from 1000 years based learning set (green) and the actual data based (red) relative frequency distributions of extreme high values.

however there is more probability weight attached to the lower and higher values compared to the actual results. In other words the relative frequency in the extreme bins is overestimated. This means that ERAP generated data has 'fatter' tails when compared to the actual distribution.

4.3.1. *Comparing ERAP statistics based on 100 years of the learning: weather-like process experiment.* In this section we reproduce all the parts of the experiment but now we use a learning set that contains 100 years. In particular, we are interested in how well the learning set represents the actual statistics, and how well the synthetic data, generated using such a reduced learning set, represents the actual statistics. The colours on all the plots are identical to the colours used in the previous experiment. All the statistics presented in this section are constructed in the same manner as for the 1000 years learning set case.

First we present the four moments on figures 22, 23, 24 and 25 respectively.

It can be seen that the mean of the learning set is almost identical to the true mean. The ERAP ensemble based relative frequency of the mean overestimates both the true and the learning set's mean, as the majority of the relative frequency is assigned to higher values of the mean by about a degree. The actual and the learning set based mean however falls into a high relative frequency bin.

In the case of standard deviation statistic the learning set based measurement and the actual measurement are identical. What we can see in

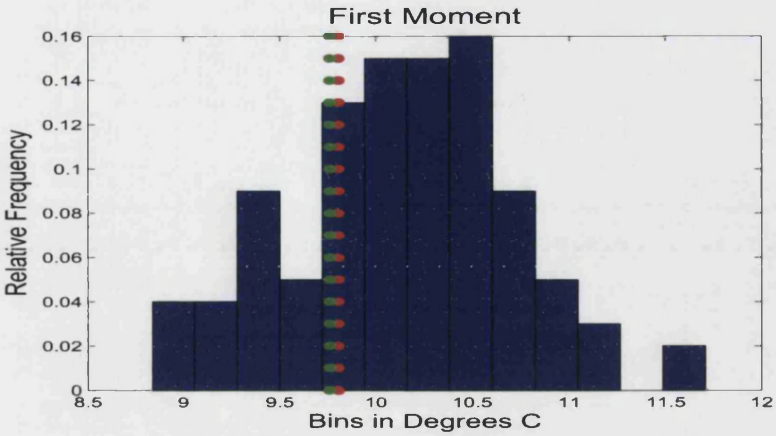


FIGURE 22. Comparison of the distribution of the first moment of the ERAP ensemble (blue) to the first moment computed from the 100 years of the learning set (green) and the actual data based mean (red).

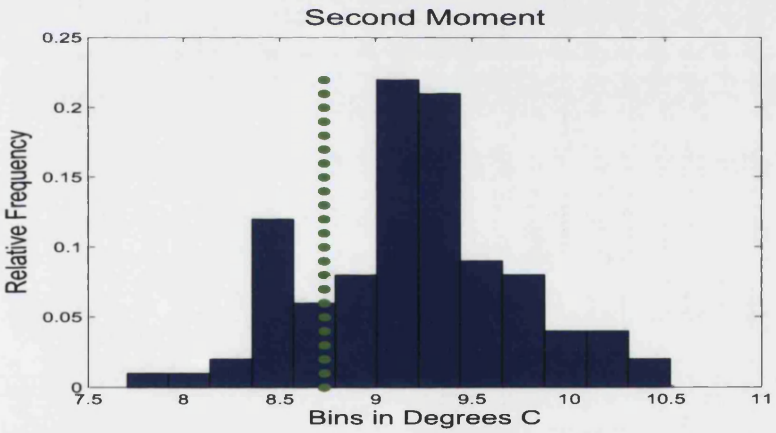


FIGURE 23. Comparison of the distribution of the standard deviation of the ERAP ensemble (blue) to the standard deviation computed from the 100 years of the learning set (green) and the actual data based (red).

the ERAP synthetic data histogram, is that the standard deviation is overestimated, and the majority of the probability weight is assigned to higher

values of the standard deviation. The relative frequency distribution is different in its characteristic when compared to the relative frequency of the standard deviation constructed using a 1000 year learning set. This shows that the statistic generated using the ERAP approach is significantly influenced by the size of the learning set. The estimation of standard deviation in case of the 100 year learning set could be improved by parameter optimisation. The first and second moments are affected by how well long term patterns are captured. We know that in the case of 100 year learning set some of the long term characteristics of the weather-like process are not well captured.

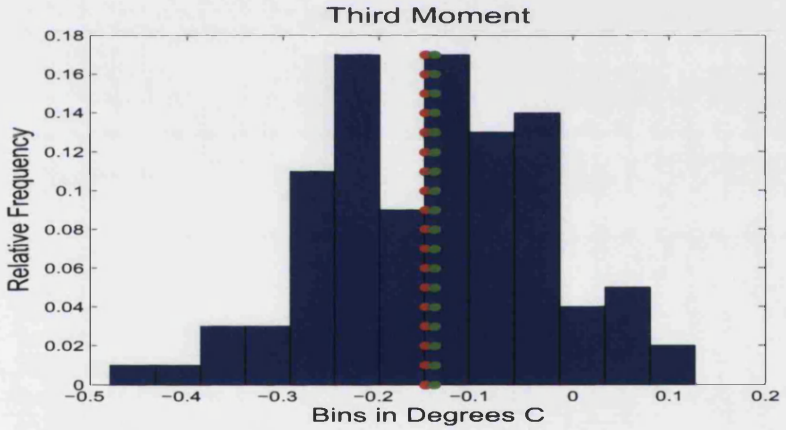


FIGURE 24. Comparison of the distribution of the third moment of the ERAP ensemble (blue) and the third moment computed from the 100 years of the learning set (green) and the 'true' third moment (red).

Actual skewness and kurtosis (figures 24 and 25 respectively) are well represented by the learning set and the ERAP synthetic data relative frequency histogram. Again the relative frequency distribution are different

compared to the distribution obtained using 1000 years of data for the learning set.

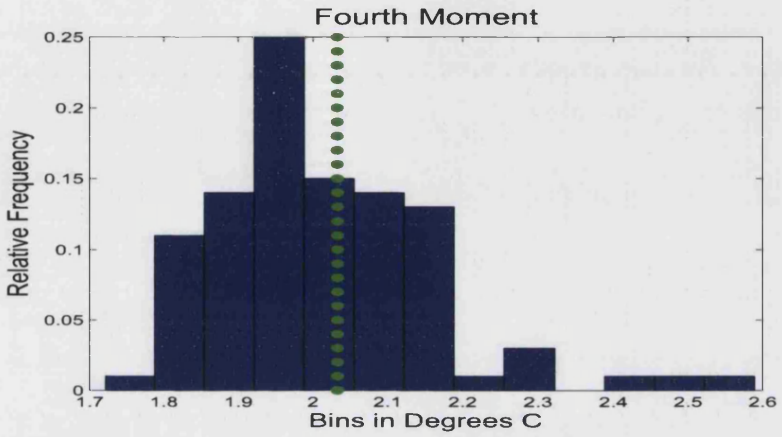


FIGURE 25. Comparison of the distribution of the fourth moment of the ERAP ensemble (blue) and the fourth moment computed from the 100 years of the learning set (green) and the actual data based (red).

Next we consider the relative frequency of all observed values, figure 26, where again blue histogram represents statistics computed from the synthetic data (data generated by the ERAP approach constructed using data from all ensemble members), green is the learning set and red the actual relative frequency histograms.

The shape of all histograms are consistent with each other, in particular, the learning set represents the relative frequency of all observed values well when compared to the actual data. The synthetic data based histogram overestimates frequency in the warmer part of the distribution and slightly underestimates the frequency in the mid range of values.

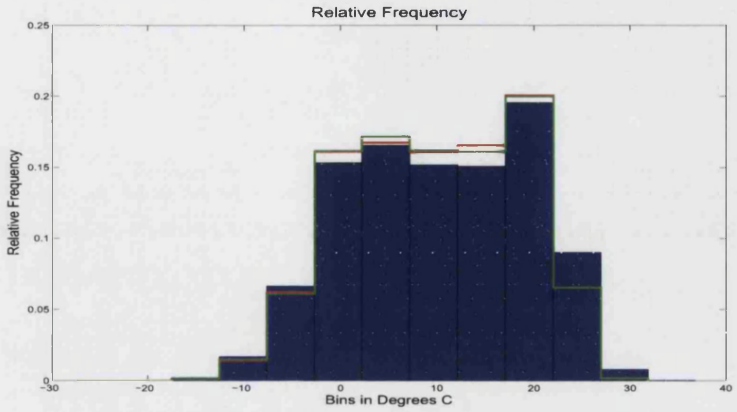


FIGURE 26. Comparison of the relative frequency distribution of all observed values in the generated ERAP ensemble (blue) to the relative frequency of the learning set (green), where learning set contains 100 years and the 'true' relative frequency.

We also, following the experimental design, examine the yearly percentiles (constructed in an identical manner to the 1000 year learning set case) and the 5th and 95th percentile histograms of the synthetic data together with the actual and the learning set measurements are plotted.

The synthetic data based yearly 95th and 5th percentiles (blue) well capture the verification data (red), figure 27 where the verification data goes outside the 95th and 5th percentile boundaries approximately 10% of the time.

Both the actual 95th (figure 28) and the 5th (figure 29) percentiles are captured well by the learning set. The actual 95th percentile statistic is reproduced better by the synthetic data compared to the 5th, as the actual 95th percentile measurement falls into a higher frequency bin.

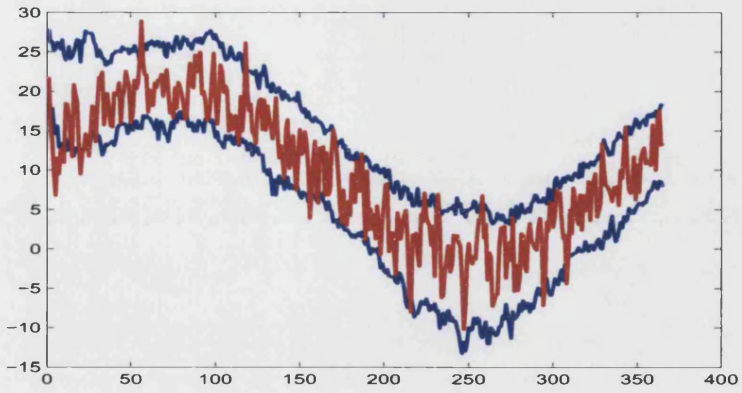


FIGURE 27. Comparison of the relative frequency of the ERAP based yearly percentiles (blue), 100 years learning set based (green) and the actual yearly percentiles (red).

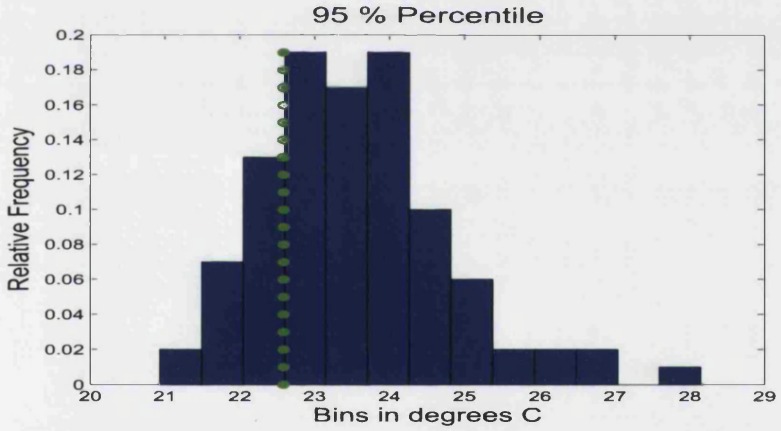


FIGURE 28. Comparison of the relative frequency of the ERAP based 95th % percentile (blue), learning set based in green, where learning set contains 100 years and the actual 95th % percentile (red).

We also consider the consecutive increasing and decreasing days. As before, as the number of consecutive days increases the observed instances

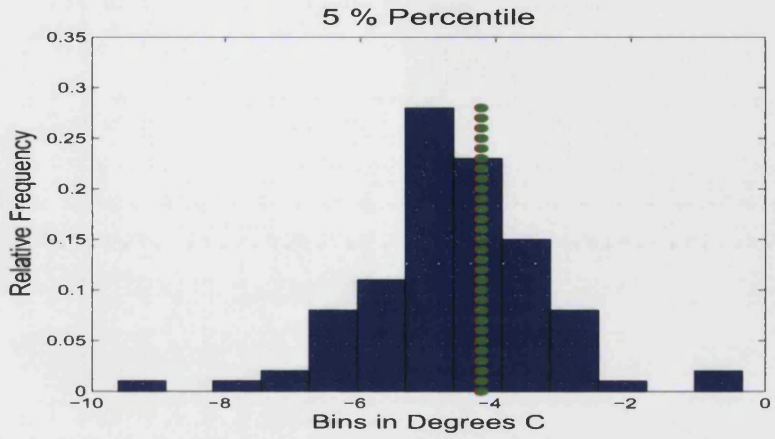


FIGURE 29. Comparison of the relative frequency of the ERAP based 5th % percentile (blue), yearly percentiles computed from 100 years based learning set (green) and the actual 5th % percentiles (red).

decrease. One can see how the size of the learning set affects this statistic in the output synthetic data histograms.

Figures 30 and 31 display the consecutive decreasing and consecutive increasing statistic respectively. The blue corresponds to the synthetic data based relative frequency distribution of the consecutive decreasing/increasing days, where the day count was computed for each ensemble member and then arranged into a relative frequency histogram. Green represents the consecutive days count based on the 100 year the learning set, red shows the actual consecutive days count and finally yellow is the mean of the synthetic data consecutive days count.

It is evident that the true one and two consecutive decreasing days statistic is represented well by both the learning set and the synthetic relative frequency distribution (as the actual measurement falls in to the maximum

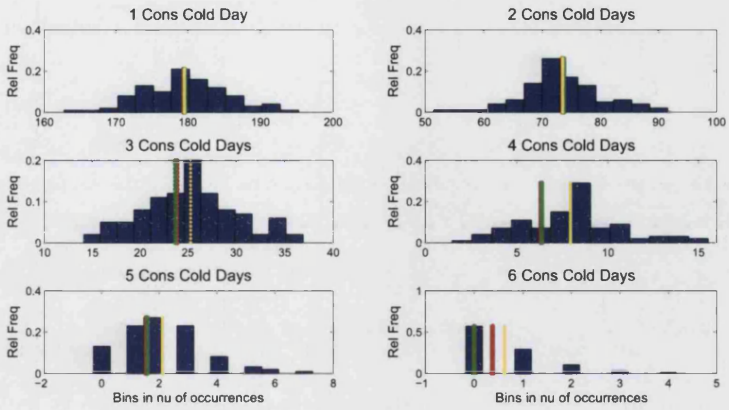


FIGURE 30. Comparison of the relative frequency of the ERAP based consecutive decreasing days (blue), together with the mean on that distribution (yellow), learning set that contains 100 years based (green) and the actual consecutive decreasing days count (red).

probability bin). The actual four consecutive cold days are not represented as well by the synthetic relative frequency histogram, as it was with the 1000 year learning set based synthetic data. For the case of 6 consecutive cold days, the learning set based measurement falls into the maximum probability bin of the synthetic relative frequency histogram, but again the number of bins must be reduced due to the limited number of occurrences. Increasing consecutive days are generally better represented by the synthetic relative frequency distribution. Overall we can observe that the 1000 year based synthetic data produced more consistent statistics compared to the true measurements.

Next we look at the distribution of the number of freezing days, again an important measurement for the purpose of weather hedging, figure 32.

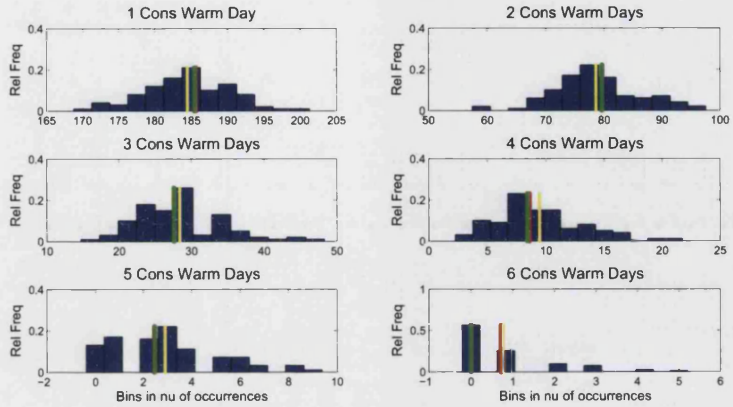


FIGURE 31. Comparison of the relative frequency of the ERAP based consecutive increasing days (blue) together with the mean on that distribution (yellow), learning set of 100 years based (green) and the actual consecutive decreasing days count (red).

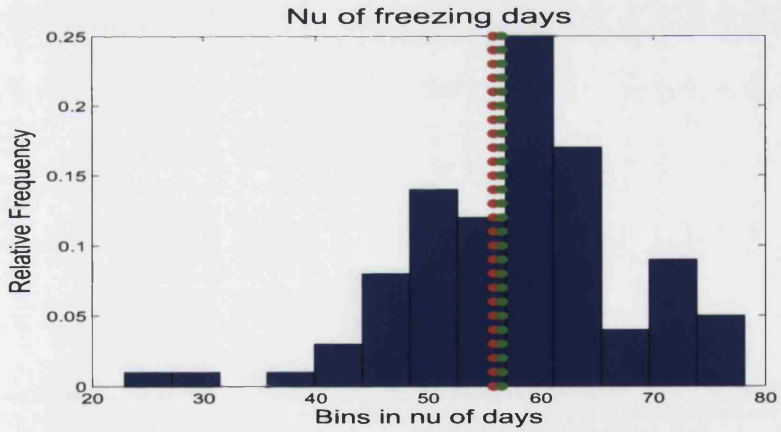


FIGURE 32. Comparison of the relative frequency of the ERAP based freezing days count (blue) to the freezing days count based on the learning set that contains 100 years (green) and the 'true' freezing days count (red).

The relative frequency histogram of the synthetic data, captures relatively well the actual number of freezing days, where both the actual count and the learning set based count fall on the edge of the maximum probability bin. The distribution itself is less spread both in terms of the maximum and minimum observed values and the standard deviation, when compared to the freezing days count generated by data synthetic using a 1000 year learning set.

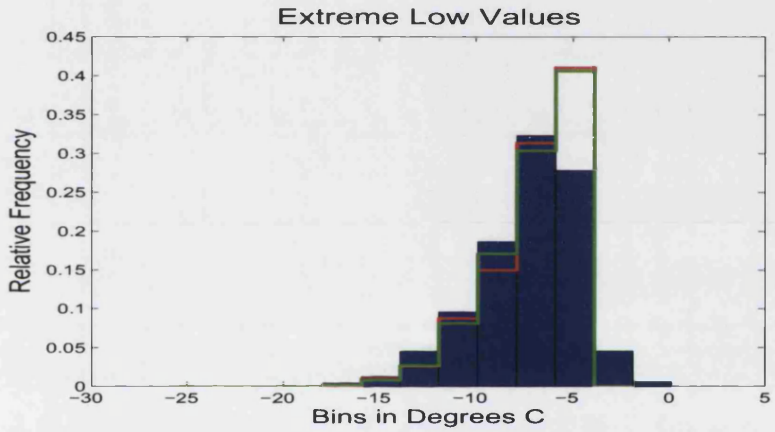


FIGURE 33. Comparison of the of the ERAP based (blue), learning set consisting of 100 years (green) and the actual data based (red) relative frequency distributions of extreme low values.

Finally figures 33 and 34 display the frequency distributions of the extreme low and and high values respectively. Again we are considering the tails of the distribution in a little more detail. One can see that the 100 year learning set statistic does not match exactly the actual distribution of the extreme low and high values. The synthetic data based distribution in this case is affected by this, particularly in the case of the extreme high

values (this is also true for the case of the extreme low values), and the relative frequencies obtained through the ERAP generated data are significantly different to both the actual and the learning set frequencies. This could be potentially improved however by parameter optimisation.

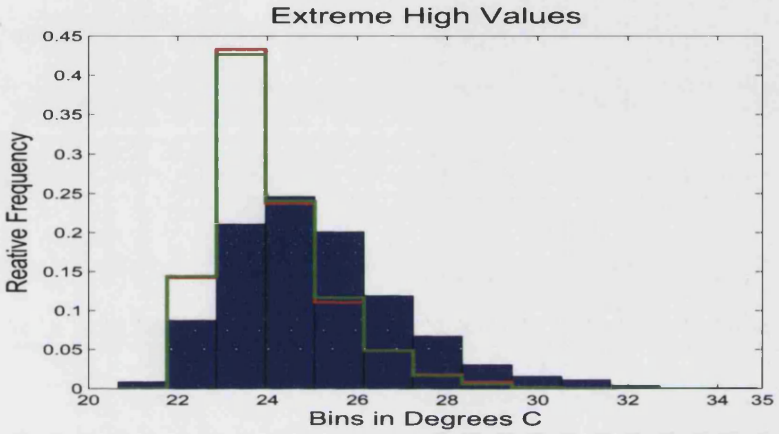


FIGURE 34. Comparison of the of the ERAP based (blue), learning set (green) and the actual data based (red) relative frequency distributions of extreme high values.

4.4. Conclusions of the ERAP controlled experiment: weather-like data. This experiment allowed us to examine how well ERAP based synthetic data reproduces the statistics of the learning set and the true statistics of the chosen weather-like system. We studied how different sizes of learning set affect such statistic estimation.

Overall we could observe that, particularly in the case of the 1000 year learning set, ERAP based synthetic weather data produced statistics that were consistent with the learning set statistics and the true statistics of the system. Furthermore, in some cases (such as 6 consecutive decreasing days) the synthetic data provided a better estimate when compared to the learning

set. In the case of such statistics as 6 consecutive decreasing days, where a limited learning set does not allow accurate estimation, it is encouraging to see that the synthetic data provides an accurate estimate of the truth.

It was also evident from the experiment that when the learning set is reduced, the quality of the synthetic data based statistic is affected. It is not necessarily, however, affected in a negative way. We observed that the freezing day count actually improved as the size of the learning set decreased.

Finally, it is important to note that the experiment was constructed with an initial choice for the parameters that were shown to be reasonable using the techniques described in section 2.1. These parameter values, however, are not optimal. As mentioned in section 2.1 the initial choices of parameters can be further optimised. It is not the intention of this thesis however to do this at this stage, and the optimisation of ERAP parameters will remain as further work.

4.5. Controlled experiment: Lorenz system. .

In this experiment we test the ERAP approach on the noisy, 'seasonal' x component of the Lorenz system [44] (see chapter 4 for details on the numerical generation of the noisy Lorenz data). In particular we generate $N_{actual} = 348,575$ sample points, with a sampling rate of 0.1 Lorenz seconds. Then noise is added, such that the noise level is approximately 2% of the size of the attractor in the x component direction. The methodology for constructing such data together with a description of the additive noise is presented in chapter 4. Finally, a sine wave is added to the noisy re-sampled x component data. The sine wave has a period of 365 data points and

amplitude equal to one. The sine wave has been added with the weather data in mind. We also choose the size of the learning set to be $N_l = 36,500$. The resulting data, that is used to test ERAP, is illustrated in figure 35.

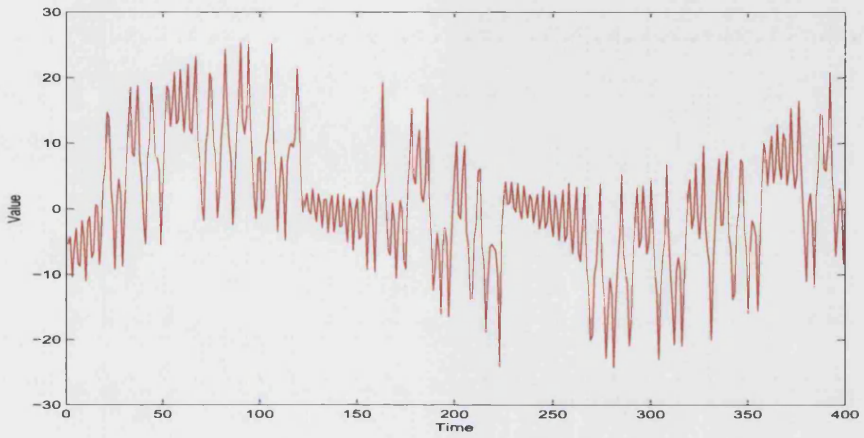


FIGURE 35. Zoom into the Lorenz noisy x component with added sine wave.

4.5.1. *Choosing parameters of ERAP for the Lorenz experiment.* Now let's consider the choice of parameters for the Lorenz experiment. It can be seen in figure 35 that the time series have regions where the values oscillate rapidly, which corresponds to a location change between the two wings of the Lorenz attractor (see picture 1). This is followed by stable, growing oscillations, which corresponds to movement on a wing of the Lorenz attractor. The sine wave shifts the Lorenz x component series such that the data appear to be 'seasonal'.

In order to determine a reasonable guess for the initial choice of the m_{long} and m_{short}, the following analysis has been applied:

- Examine the first difference for the step before and the step after a sign change.

- Record number of points between sign change. This is the area of the attractor where the movement jumps from one wing to another.
- The 90th percentile of these lengths is used as the short-term scale parameter.
- Once the length of the short term scale has been determined, we estimate the number of points between short-term spikes, and choose a large enough percentile of those lengths so that this pattern will be well captured.

The number of nearest neighbours (long and short term) and the size of the ensemble, were chosen to be the same for all the experiments: the Lorenz experiment, the weather-like process experiment and for the actual Berlin data. The Talagrand diagram was applied to confirm that the number of short term neighbours $N^{short}_{Neighbours}$ (and hence the number of the long term neighbours, as described in section 2.1) is appropriate for the Lorenz experiment. A summary of the parameters that were chosen for the ERAP generator for Lorenz data is given in table 4.5.1.

4.6. Lorenz experiment: comparing statistics. Once the ERAP parameters have been chosen for the Lorenz experiment, the ERAP generator was used to generate a synthetic data ensemble. The resulting synthetic data is displayed as individual ensemble members in figure 36, where the individual ensemble members are blue and fragments of the actual data are shown in red. In this experiment we also used a phase controlling parameter. When the phase of the chosen nearest neighbours is restricted, the resulting ERAP ensemble is more stable. This allows us to produce synthetic data

N_l	100*365
τ	1
m_{long}	41
m_{short}	20
$m_{reduced\_long}$	4
$m_{reduced\_short}$	3
$N^{long}_{Neighbours}$	28
$No^{short}_{Neighbours}$	32
$Max_{extention}$	27
$Min_{extention}$	9
$\alpha_{iteration}$	100
$Size\_of\_ensemble$	100
$seed$	69
$phase\_control\_parameter$	9

TABLE 4. Parameters of ERAP for the Lorenz experiment.

that extends further in time. In this experiment we generated 7 years of synthetic data.

Figure 37 illustrates the complete ERAP synthetic ensemble for the Lorenz experiment, together with the verification.

In this section the steps of the experimental design are followed, in order to assess the performance of the ERAP generator. First traditional statistics are compared. Figure 38 illustrates the frequency distribution of the mean generated by the ERAP ensemble (blue) together with the mean of the

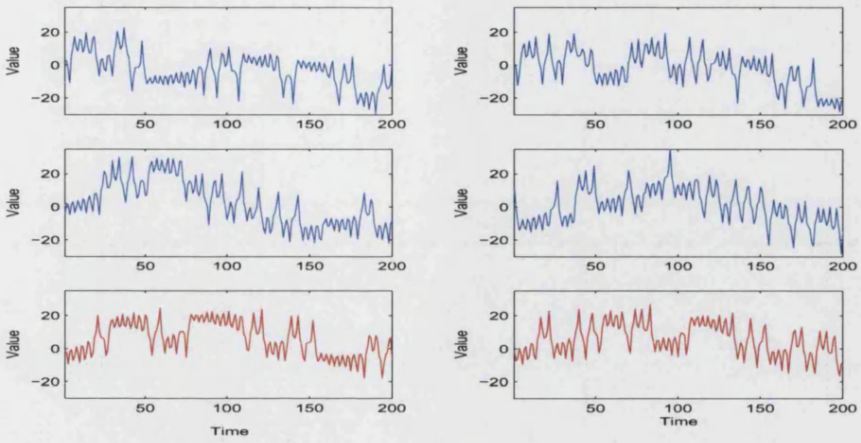


FIGURE 36. First two rows (blue) illustrate individual ensembles generated by the implemented ERAP approach. The verification data is displayed in red.

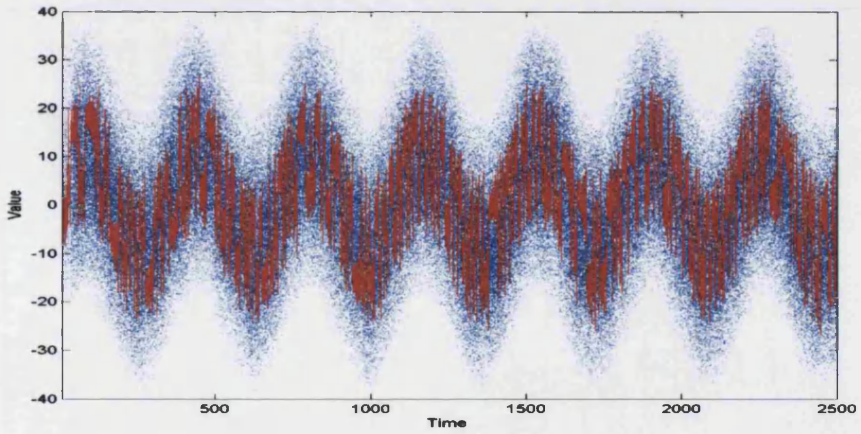


FIGURE 37. ERAP ensemble produced (blue) using the learning set, with the verification, which is given in red.

learning set (green) and the 'true' mean computed form the actual data using N_{actual} data points (red dashed line). This plot illustrates where, in this distribution the actual mean and the mean of the learning set fall. It can be seen that the ERAP based distribution overestimates the value of the actual mean, as most of the distribution is observed for higher values of

the mean. It it also evident that the learning set mean is also overestimating the actual value.

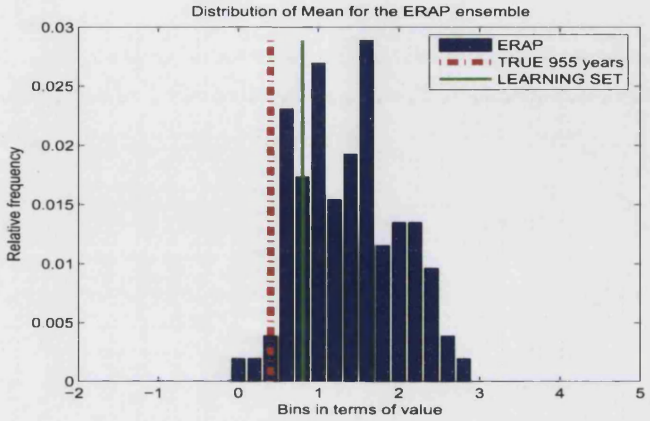


FIGURE 38. Comparison of the relative frequency distribution of the ERAP based mean (blue) together with the mean of the learning set (green) and the actual (red dashed line).

Figure 39 shows the frequency distribution of the standard deviation of the ERAP ensemble (blue) together with the standard deviation of the learning set (green) and the actual (red dashed line). This plot illustrates where, in the ERAP standard deviation frequency distribution, the actual standard deviation and the standard deviation based on the learning set fall. It can be observed that the ERAP based distribution captures both the actual standard deviation and the standard deviation of the learning set well, i.e. they both fall in the largest probability bins. It can also be observed that the standard deviation of the learning set underestimating the actual standard deviation.

Figure 40 demonstrates the relative frequency distribution of the kurtosis for the ERAP ensemble (blue) together with the kurtosis of the learning set

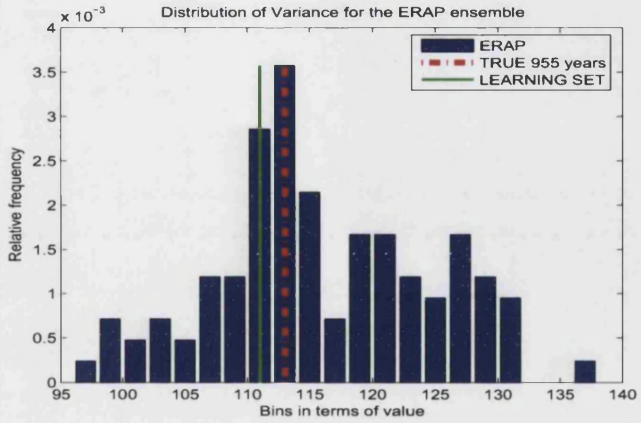


FIGURE 39. Comparison of the relative frequency distribution of the standard deviation for the ERAP ensemble (blue) together with the standard deviation of the learning set (green) and the actual data (red dashed line).

(green) and the actual kurtosis (red dashed line). It can be seen that both the actual kurtosis, and the kurtosis of the learning set are identical and well captured by the distribution based on the ERAP generated data, although they do not fall into the largest frequency bin.

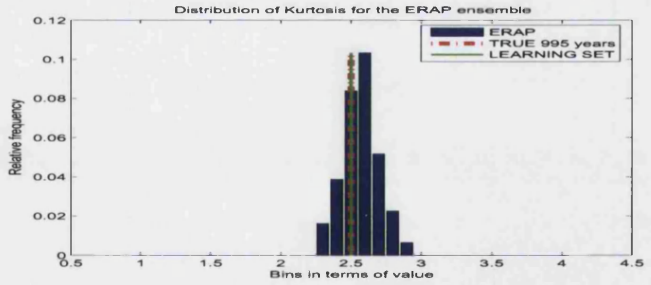


FIGURE 40. Comparison of the relative frequency distribution of the ERAP ensemble based kurtosis (blue) to the kurtosis of the learning set (green) and the actual data based kurtosis (red dashed line).

Finally, figure 41 illustrates skewness. In particular, the frequency distribution of the skewness computed from the ERAP ensemble (blue) together with the skewness of the learning set (green) and the actual skewness (dashed red line) is shown. Actual skewness and the skewness of the learning set are well captured by the ERAP based generated data and are identical to each other.

40

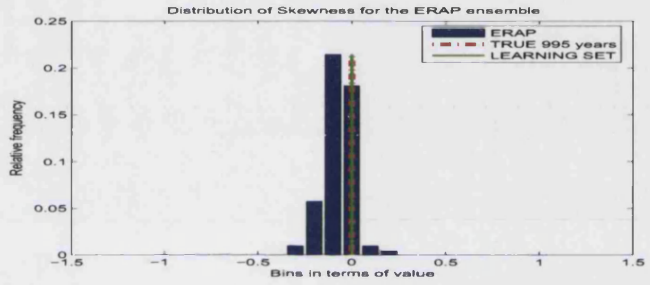


FIGURE 41. Comparison of the relative frequency distribution of the ERAP ensemble based skewness (blue) together with the skewness of the learning set (green) and the skewness based on the actual data (dashed red line).

Next the climatology, in particular the relative frequency of all observed values, is examined. Figure 42 illustrates the climatology of the actual (red), learning set (green) and the ERAP ensemble based 5th and 95th isopleths. It can be seen that the actual climatology and the climatology of the learning set are both close to each other and are captured within 90 % of the ERAP based climatology distribution.

We comparing distributions of the proportion of switches from one wing to another on the attractor. A distribution of the logarithm of base two of

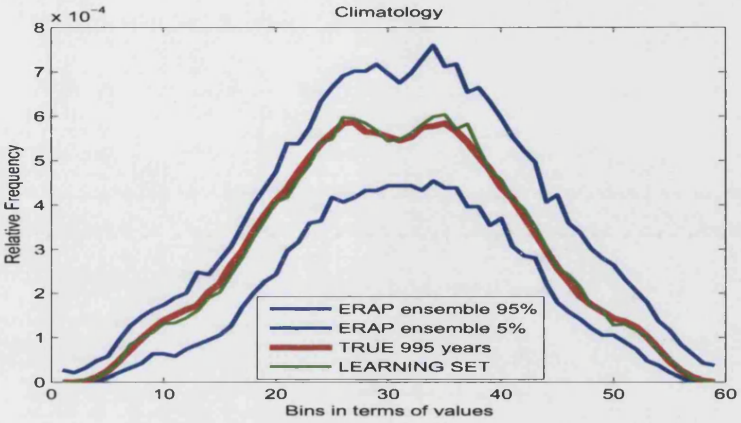


FIGURE 42. Climatology of the actual data (red), the learning set (green) and the ERAP ensemble based 5 and 95 isopleths. The isopleths were found by calculating climatology for each ensemble member.

the relative frequency of the time spend on the 'negative' wing of Lorenz attractor for the learning set shown in figure 43. It can be observed that short stay on the wing is most prominent.

Now lets compare frequency distributions of decreasing/increasing consecutive days. The relative frequency distribution of the number of consecutive unique decreasing data points for the 5th, 50th and 95th isopleths of the ERAP ensemble (blue), the learning set (green stars) and the actual (red dashed line) is shown on figure 44.

Similarly the relative frequency distribution of the number of the consecutive unique increasing data points for the 5th, 50th and 95th isopleths of the ERAP ensemble (blue), the learning set (green stars) and the actual (red dashed line) is illustrated on figure 45.

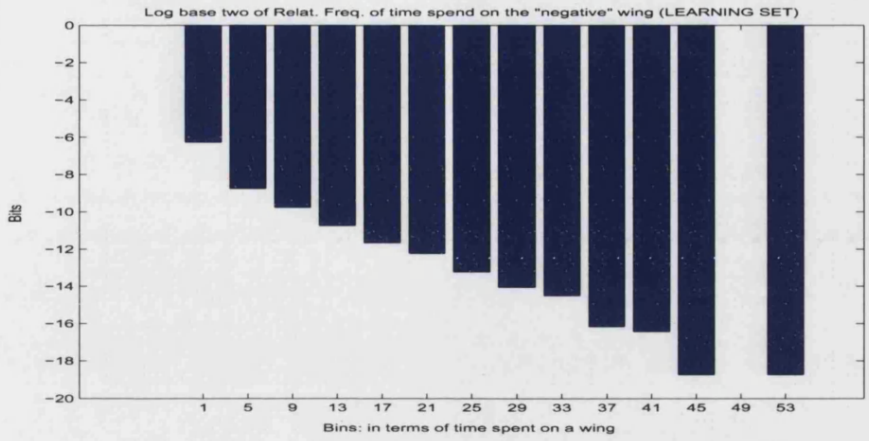


FIGURE 43. A distribution of the logarithm of base two of the relative frequency of the time spend on the 'negative' wing of Lorenz attractor for the learning Set. It can be observed that short stay on the wing is most prominent.

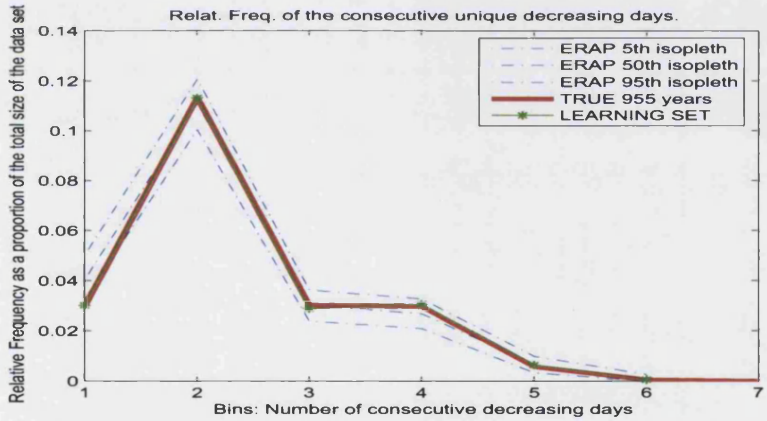


FIGURE 44. Relative frequency distribution of the number of the consecutive unique decreasing data points for the 5th, 50th and 95th isopleths of the ERAP ensemble (blue), the learning set (green stars) and the actual (red dashed line).

It can be noted that for both, the consecutive decreasing and increasing days the learning set's based statistic and the actual are both well captured by the ERAP produced ensemble.

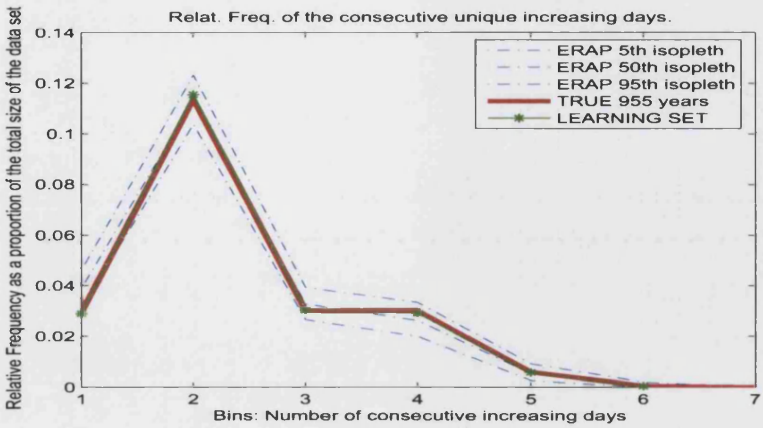


FIGURE 45. Relative frequency distribution of the number of the consecutive unique increasing data points for the 5th, 50th and 95th isopleths of the ERAP ensemble (blue), the learning set (green stars) and the actual (red dashed line).

4.7. Conclusions of the ERAP controlled experiment: Lorenz system. As in the case of the weather-like experiment the true statistical properties of the Lorenz noisy re-sampled x component (with added sine wave) data is represented well by the synthetic data generated using the ERAP approach.

Particularly in the case of standard deviation estimation, it was observed that the synthetic data based frequency captured the true standard deviation well, and the actual standard deviation fell into the maximum probability bin. In that particular case, the learning set standard deviation was significantly different to the true standard deviation. The percentiles were also well represented, both in terms of the relative frequency distribution and the actual comparison.

Again the optimisation of the ERAP parameters for the Lorenz case might potentially improve the skill of the synthetic data, and its ability to reproduction key statistical measurements.

5. ERAP experiment on the real weather data: Berlin max daily temperature.

In the last section of this chapter synthetic weather data is generated for the Berlin daily max temperature using ERAP. The Berlin data that is used as a learning set (125 years) has been described in chapter 3. In this section the actual observed values of one year that occurred after the last point in the leaning set are referred to as the verification or the actual (the verification set contains: $N_l + 1$,...,$N_l + 365$ data points).

The parameters that were chosen for the test weather process were also used for real weather data. Example singular values and vectors are presented in figures 46 and 47 respectively. They confirm that the initial choice for $m_{long}^{reduced}$ is reasonable, according to the methodology given in section 2.1. The other parameters are also examined using the methodology of section 3.

After the initial choice of parameters had been determined, the ERAP synthetic weather generator was applied to simulate the Berlin synthetic daily maximum temperatures that are used in chapter 6 to price a weather derivative. The resulting ensemble, and the verification is presented in figure 48.

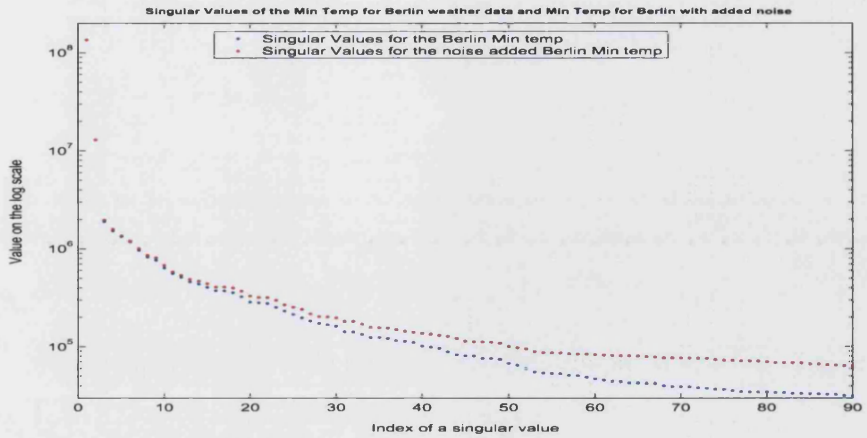


FIGURE 46. Singular Values for the Berlin max temperature and singular values for the noise added Berlin max temperature for the long term signal

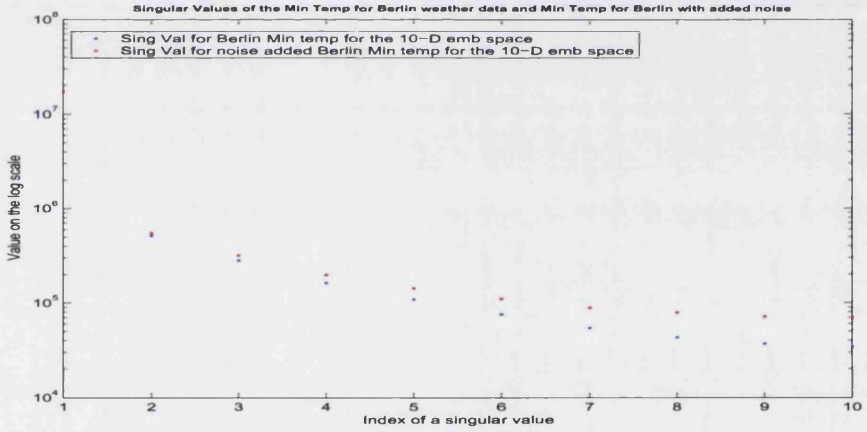


FIGURE 47. Singular Values for the Berlin Min temperature and singular values for the noise added Berlin max Temperature for the short term signal.

It can be seen that the generated ensemble data looks reasonable. In order to examine whether the generated synthetic data represents the verification and the learning set well, we compute all the statistics described in section 3 following the steps of the experimental design. It is of particular

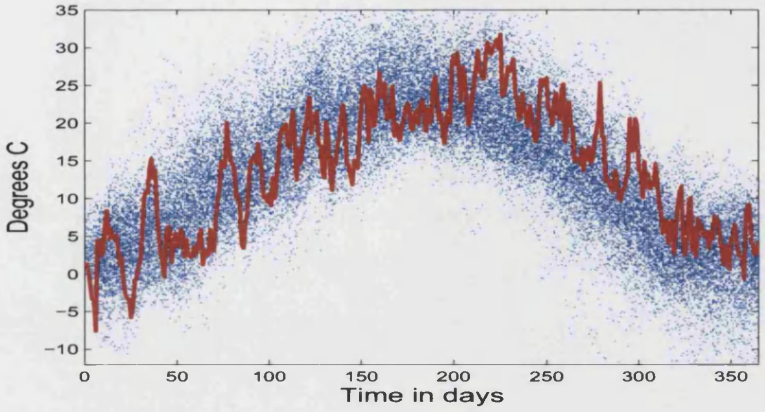


FIGURE 48. ERAP ensemble (blue) and the verification (red) for the Berlin data.

interest how the learning set statistics vary when compared to the verification data statistics and where the synthetic data statistics provide better insight into the characteristics of the Berlin data, beyond the capabilities of the learning set alone.

5.1. Comparing statistics of the ERAP ensemble with the statistics of the historical observations - Berlin daily maximum data. First we examine the relative distributions of moments of the synthetic data, the moments of the learning set and the moments of the verification. The frequency distribution of the moments of the synthetic data, as before, are represented in blue in all the plots in this section. The statistical measurements computed from the learning set are presented in green, and finally the verification based measurements are given in red.

Figures 49, 50, 51 and 52 show first, second, third and fourth moments respectively. In the case of the first moment the learning set overestimates the mean of the verification by approximately one degree. The distribution

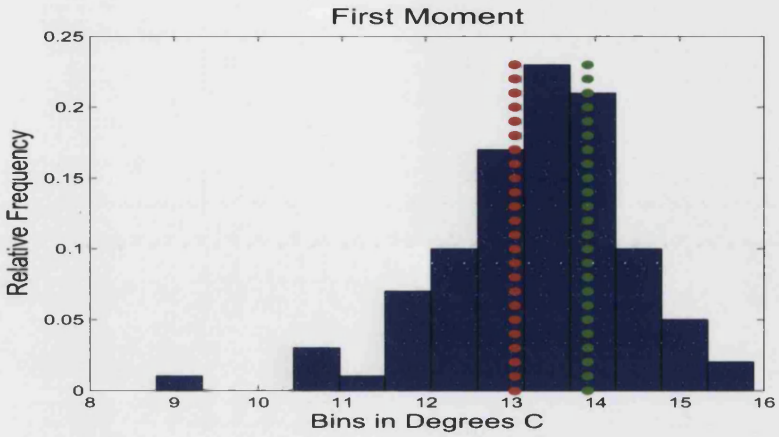


FIGURE 49. Comparison of the distribution of the first moment of the ERAP ensemble (blue) to the first moment of the learning set (green) and the verification based mean (red) for the Berlin data.

of the mean of the synthetic data, however, the maximum probability bin falls between the values of the verification mean and the learning set's mean. In the case of the mean, the synthetic data is both consistent with the learning set and offers enhanced statistics, as the mean of the verification set is well captured.

In the case of the standard deviation, the learning set standard deviation is not identical to the verification set standard deviation. The relative frequency distribution of standard deviation constructed using the ERAP generated data captured the actual standard deviation well, since the actual standard deviation fall on the edge of the maximum probability bin. A similar situation occurs in the case of the skewness.

The fourth moment computed from the verification does not fall into the maximum probability bin, whilst the learning set estimation of the fourth

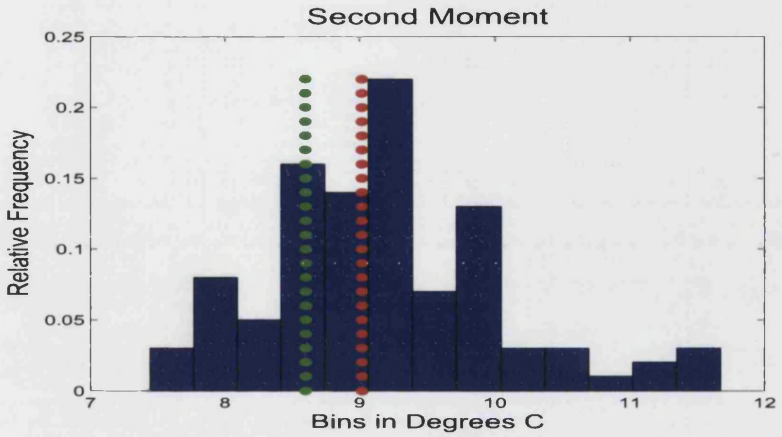


FIGURE 50. Comparison of the distribution of the standard deviation of the ERAP ensemble (blue) to the standard deviation of the learning set (green) and the verification based standard deviation (red) for the Berlin data.

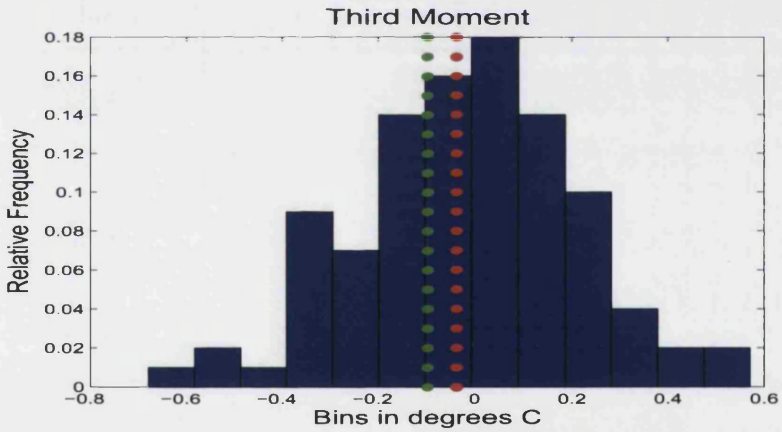


FIGURE 51. Comparison of the distribution of the third moment of the ERAP ensemble (blue) to the third moment of the learning set (green) and the verification based third moment (red) for the Berlin experiment.

moment does. It is evident, however, that the ERAP ensembles fourth moment distribution contains bins with positive frequency of kurtosis values

that are above the verification based estimate. This is particularly relevant, as the learning set kurtosis underestimates the verification kurtosis, figure 52.

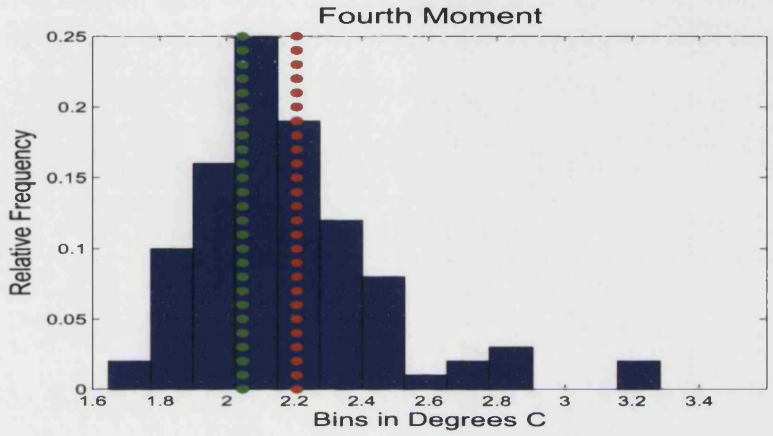
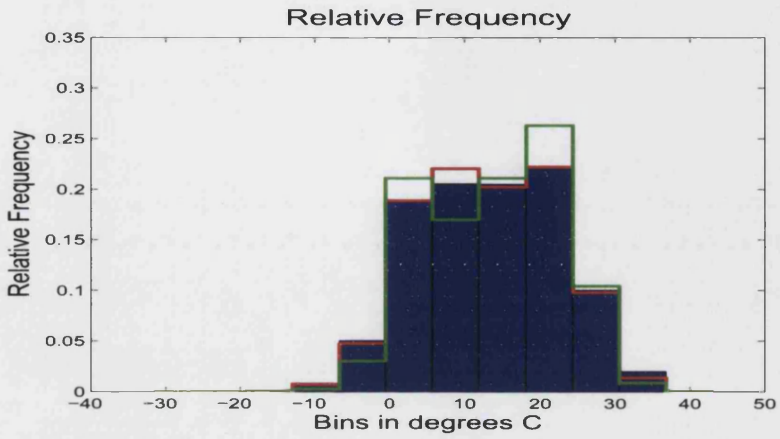


FIGURE 52. Comparison of the distribution of the fourth moment of the ERAP ensemble (blue) to the fourth moment of the learning set (green) and the verification based fourth moment (red) for the Berlin experiment.

Next we consider the relative frequency distribution of all observed values, give in figure 5.1. It is evident from the plot that the ERAP relative frequency follows the verification relative frequency more closely when compared to the learning set relative frequency.

When we looked at the yearly 5th and 95th percentiles of the ERAP synthetic data, it was evident that ERAP generated data captures well the verification series, particularly in the first half of the generated year. It can be observed that with real weather data, the skill of the generated synthetic series reduces as generated time increases, figure 53. This is a particulary interesting example of performance of the ERAP, in particular



Comparison of the rel freq. distr of ERAP ensemble, actual and the learning set: Berlin data.] Comparison of the relative frequency distribution of the of the ERAP ensemble (blue) to the relative frequency distribution of the learning set (green) and the verification based relative frequency distribution (red).

its limitations. The second half of that particular verification year was the hottest on record in Germany, and hence is not well captured by the ERAP generated ensemble. This is because the ERAP ensemble will only contain measurements that came from the historical data.

Next, the 5th and 95th percentiles of the verification set and the learning set were compared to each other, and the frequency histogram of the percentiles of the ERAP synthetic Berlin weather data is shown in figures 54 and 55 respectively.

The 95th percentile of the learning set is almost identical to the 95th percent percentile of the verification. And both measurements are well represented by the synthetic relative frequency histogram of the 95th percentile, as both measurements fall into the maximum probability bin. In the case

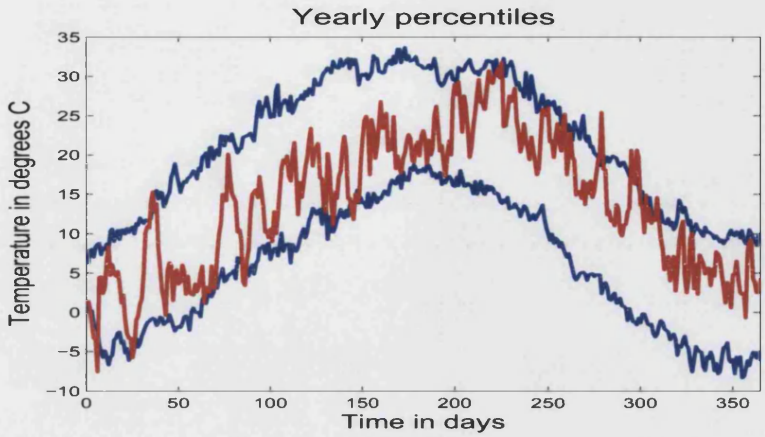


FIGURE 53. Comparison of the ERAP based yearly 95th and 5th percentiles (blue) and verification yearly data (red) for the Berlin experiment.

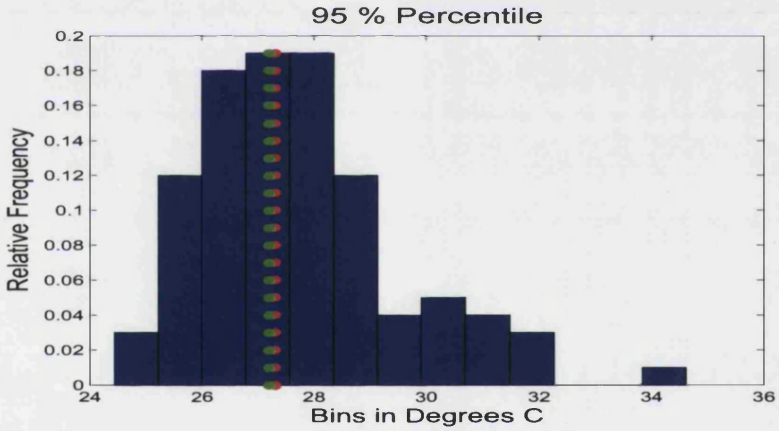


FIGURE 54. Comparison of the relative frequency of the ERAP 95th % percentile (blue), 95th % percentile computed from the learning set (green) and the 95th % percentile of the verification (red) for the Berlin experiment.

of the 5th percentile, it is quite a different story. The learning set 5th percentile is different to the verification 5th percentile by around 2 degrees. The verification 5th percentile falls on the edge of the maximum probability bin

of the ERAP synthetic data histogram of 5th percentiles. The distribution itself in this case exhibits large uncertainty, figure 55.

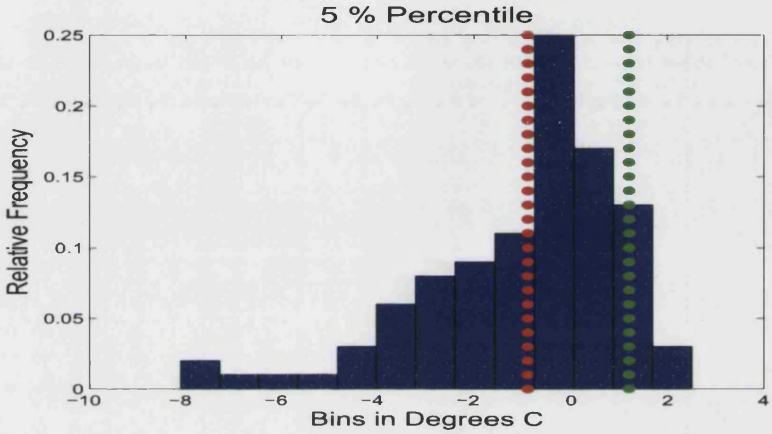


FIGURE 55. Comparison of the relative frequency of the ERAP based 5th % percentile (blue), 5th % percentile computed from the learning set (green) and the 5th % percentile of the verification (red).

In the case of the consecutive increasing and decreasing days count that are presented in figures 56 and 57 respectively, it can be observed that in most cases, the mean of the ERAP frequency distribution of the days count, is almost identical to the count of the verification set (apart from 2 consecutive days case). In comparison, the learning set count does not estimate the verification statistic as well. This is even more prominent in the case of the increasing days count.

The freezing days count statistic was computed for the verification set and the learning set, both measurements were compared to the ERAP synthetic data frequency histogram (figure 58). The learning set underestimated the number of freezing days of the verification data by about 10

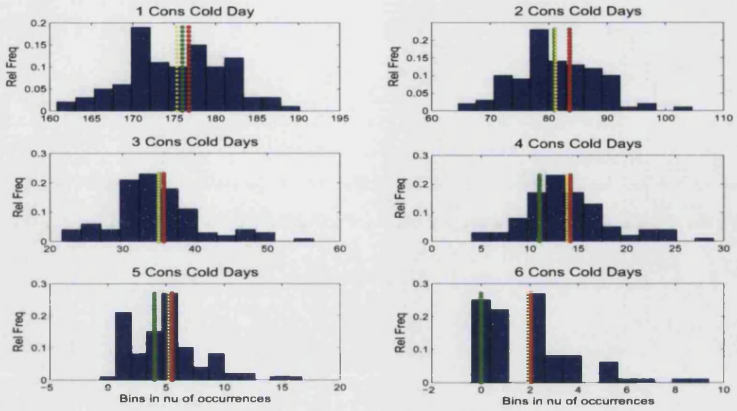


FIGURE 56. Comparison of the relative frequency of the ERAP consecutive decreasing days (blue), together with the mean of that distribution (yellow), learning set (green) and the verification consecutive decreasing days count (red).

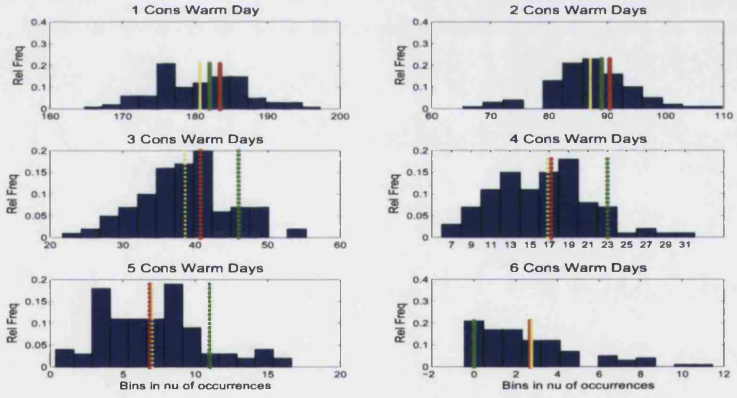


FIGURE 57. Comparison of the relative frequency of the ERAP consecutive increasing days (blue), together with the mean of that distribution (yellow), learning set (green) and the verification consecutive increasing days count (red).

instances. And although the learning set based measurements fell into the maximum probability bin of the ERAP based distribution, the verification

based statistic falls in the the second largest probability bin. Additionally it can be observed that there are significant positive frequencies above the verification value that were generated by the ERAP ensemble for the freezing day count statistic.

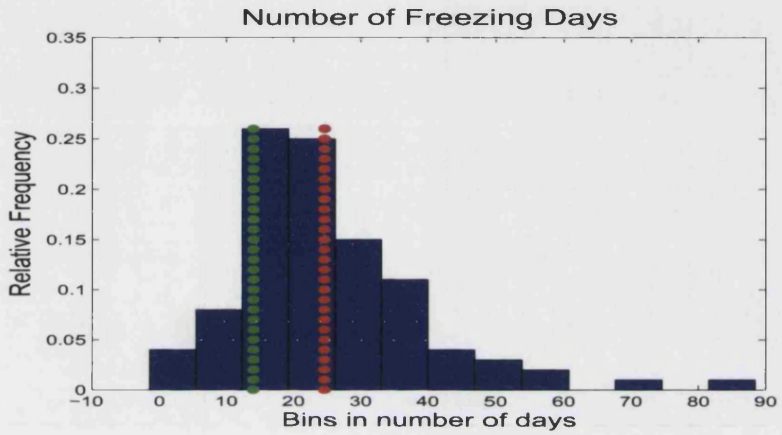


FIGURE 58. Comparison of the relative frequency of the ERAP based freezing days count (blue) to the freezing days count based on the learning set (green) and the verification based freezing days count (red) for the Berlin data.

Finally the relative frequency distributions of the extreme low and high values were compared for the verification data (red), the learning set (green) and the synthetic data ensemble (blue). These statistics are presented in figures 59 and 60 respectively.

It can be observed especially in the case of the extreme low value distribution, that the verification distribution is better represented by the ERAP ensemble distribution than the learning set distribution. That is because the difference in the relative frequencies observed between the verification and the ERAP distributions are smaller in magnitude than the difference

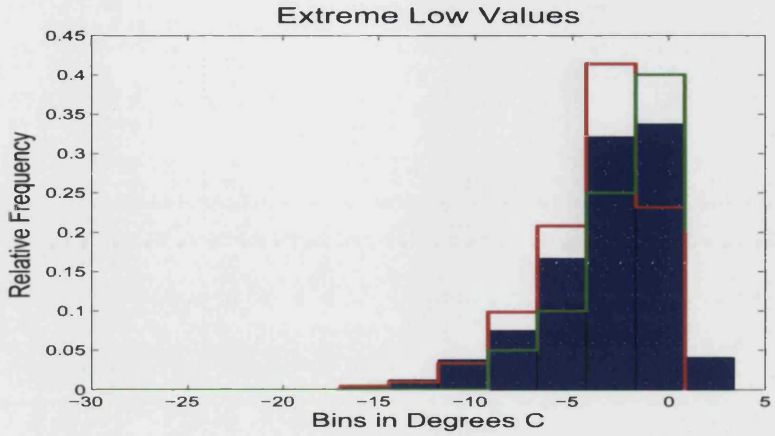


FIGURE 59. Comparison of the of the ERAP based (blue), learning set based (green) and the verification data based (red) relative frequency distributions of the extreme low values for the Berlin data.

between the frequencies of the verification and the learning set, across all the bins.

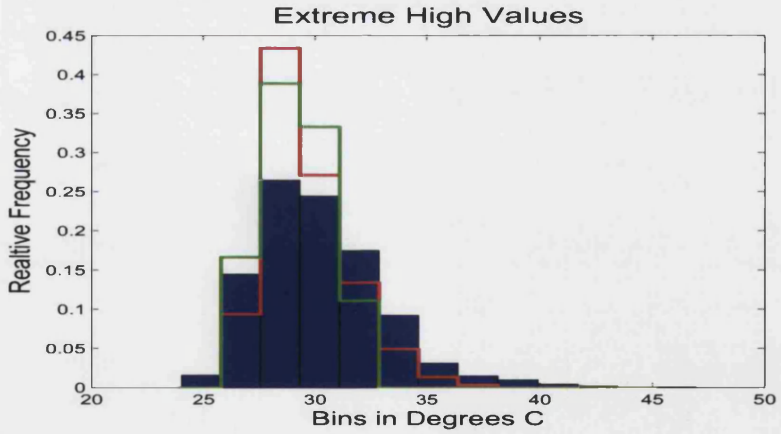


FIGURE 60. Comparison of the of the ERAP (blue), learning set (green) and the verification data (red) relative frequency distributions of the extreme high values for the Berlin data.

Summarising the results of this section, it can be concluded that all the verification based statistical measurements, described in the experimental design, are well and in some cases better represented by the synthetic ERAP than by the learning set. This could be further improved through parameter optimisation.

CHAPTER 6

Implications for pricing and predicting.

1. Overview

In this chapter, the different approaches to producing future temperature scenarios, which have been used to price weather derivatives (discussed in chapters 2 and 3) are compared to ERAP based pricing (chapter 5). All the methods that were used generated future temperature scenarios when physical weather forecasts were not available.

The comparison is made between the prices obtained using:

(1) The historical data based actuarial method, which uses samples of historical data. Also in this category we examine pricing using fitted distributions. In particular, a Normal distribution fitted to the equivalent historical data sample. Both pricing methodologies follow Zeng's [**29**] paper, adapted to our option specification.

(2) The statistical temperature modelling approach based on, Cao and Wei's paper [**30**] and pricing using Davis' paper [**31**] [1];

(3) Finally, the ERAP approach with all the historical data available as an input to the ERAP re-sampler.

[1] using 20 years for calibration

All the obtained prices are compared, together with additional characteristics of the Cooling Degree Day scenarios, such as: standard deviations of option payoffs and Value At Risk (VaR).

The weather option that is priced in this section is: a July Cooling Degree Day(CDD) Call, based on daily maximum observations. In this case the daily maximum observations are used, rather than the standard daily averages that are often used as a basis of the CDD calculation (see section 2 for the definition on CDD)[2]. The option is written for Berlin, with a strike of $K = 130$ CDD converted to currency at the rate $k = 100£$ per Degree Day.

The last section of this chapter briefly discusses how the information from weather forecasts could be added to the ERAP generated temperature simulations, using the methodology described in chapter 4, in order to produce more 'accurate' prices.

The data used was 126 years of Berlin daily maximum temperatures (01/01/1876-31/12/2002 inclusive, see chapter 3 for full details). As the verification set, 3 years were used from 2003 to 2005 inclusive, figure 1. Three years were used to compose the verification set, as the daily maximum temperature observations in July of each of those three years varied extensively.

It can be seen that in 2003, the beginning of July experienced a heat wave and the overall daily maximum temperature stayed at a very high level. 2004 had the most stable July in the verification set and the daily

[2]This is due to the data that was available

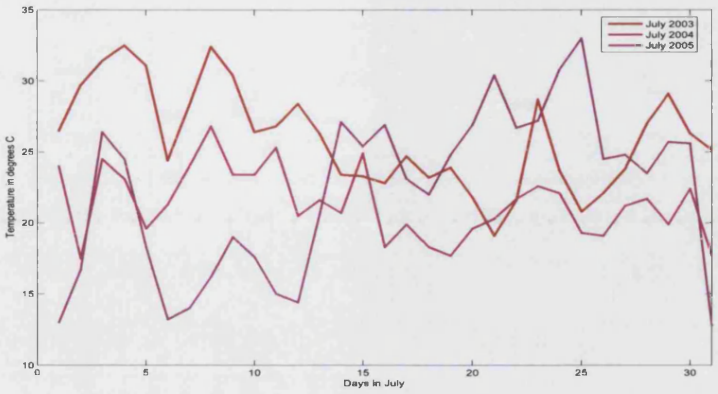


FIGURE 1. Daily max temperature observations for July for the period of 2003-2005, Berlin. Red-2003; pink - 2004; purple - 2005.

maximum values were close to the 'expected' level. July 2005 experienced rather low maximum values in the beginning of the month, followed by a rapid rise. We want to see a variety of payoffs, in order to see what happens for different temperature behaviour.

2. Pricing the CDD July Call using historical data only.

We begin by adapting the steps taken by Zeng [**29**] and consider a different number of years as a basis for the price of the CDD Call. We extend the sample used by Zeng, and also consider the price obtained with 126 and 2 most recent years of the historical data. The 126 year based price is useful, as we also used 126 most recent years as an input to the ERAP approach. The motivation behind the 2 year based price, is the tendency in the financial industry to use as little historical data as possible. It is feared that additional years might result in underpricing due to recent Global warming temperature shifts.

In section 2.1 each historical year is considered as a temperature scenario. In section 2.2 temperature scenarios are drawn from a Normal distribution fitted to the historical data.

For each scenario, for a chosen month, the CDD value is computed, using equation 1, with A_i being the daily maximum temperature (instead of an average daily temperature) and T is the number of days in that particular month. By the definition of the CDD Call, if the monthly CDD count is above the strike level, the difference between the monthly CDD and the strike is multiplied by the price per CDD. This calculation defines the payoff of the option (see chapter 2) for each temperature scenario. Finally, to determine the payoff for the Berlin CDD Call, an average of the payoffs over all scenarios is taken (see chapter 2 on the actuarial pricing methodology).

2.1. Comparing characteristics of CDD produced using historical data only. July of each historical year is considered as a scenario for pricing. Figure 2 illustrates the CDD count and the corresponding payoffs that would have occurred in each year for the full historical record (126 years: 1876 - 2002 inclusive).

As the option payoff is computed as an average over several years, various subsets of the 126 most recent years are considered, following Zeng's methodology [**29**]. It can be seen, that in the earlier years, the CDD count, and hence, the payoffs exhibit multiple upward spikes. Overall, the data at the beginning of the historical record displays higher maximum temperature measurements for July. This would result in a higher average payoff if earlier

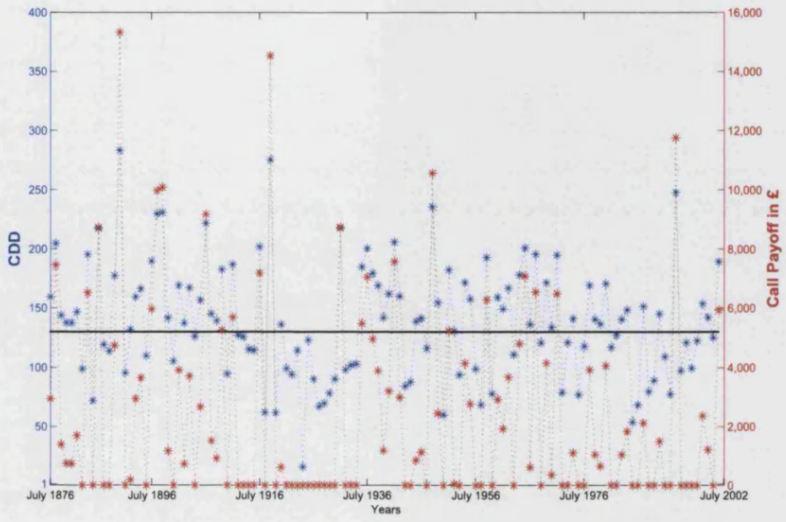


FIGURE 2. Cooling Degree Days and Payoffs of the CDD July Call for each historical year: 1876-2002. Blue stars are the monthly CDD recorded for July of each historic year. Red stars are the corresponding payoffs in £.

years are included. Figure 3 shows payoffs, standard deviations of payoffs and VaR computed using different number of years.

The CDD Call prices (payoffs) for Berlin, based on different number of years, together with the standard deviation and VaR are also summarised in table 1. The payoffs, as concluded in Zeng's paper, increase as the size of the most recent historical record decreases. This is only true, however, starting from the 30 year record. Both, the 51 year record and the full 126 year record, produced payoffs at a higher level. The payoff based on the 126 years of data is higher than the payoff based on 10 years. This finding does not contradict Zeng's conclusions, at least with regards to the 51 year record. He was dealing with daily average temperatures, rather than maximum in

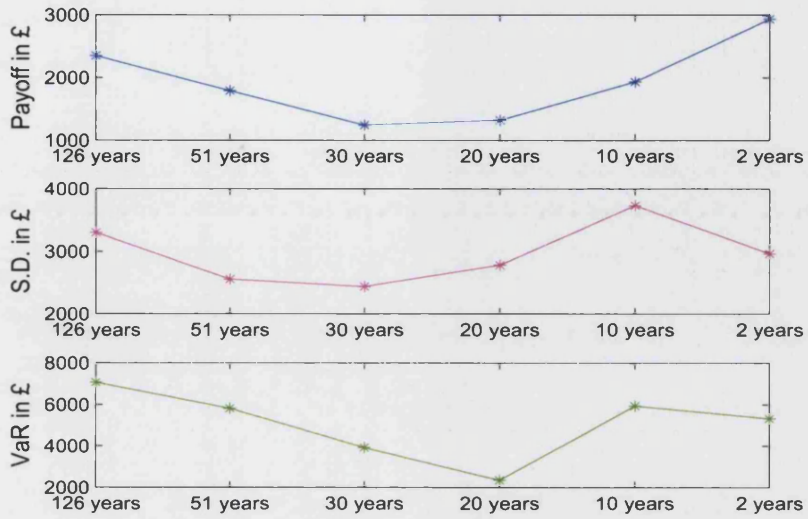


FIGURE 3. Payoffs, standard deviations and VaR measurements for the July CDD Call, computed using different number of years. Blue stars are the payoffs in £; pink stars are the standard deviation of the payoff in £; finally the green stars are the VaR also in £

this case, and additionally the CDD option in Zeng's paper was priced for the US, rather than Europe.

Now, let's examine the payoff that occurred in the verification years. In other words, if the option would have been priced using the actuarial historical data based price, how many years, as the basis of pricing, should have been considered. Also, how 'wrong' was the price. The payoff of July 2003, 2004 and 2005 are considered together with the average for those three years. Additionally, boundaries are created around the payoffs according to the standard deviations, to see whether the actual payoff would have been captured if the standard deviation had been added to the price. This is a

Years included in calculation	Mean Payout	S.D. of Payout	VaR
126	2349.69	3299.12	7062
51	1790.96	2555.59	5816.00
30	1242.58	2438.30	3920.00 20
20	1320.95	2783.84	2360.00
10	1932.73	3733.72	5930.00
2	2935.00	2965.46	5319

TABLE 1. Payoffs, s.d. and VaR for the CDD July Call for Berlin, computed suing different number of years samples.

useful assessment of performance, as the price of a weather contract could be defined as average payoff plus or minus one standard deviation, depending upon whether the user is sensitive to under or over pricing.

Figure 4 illustrates the payoffs computed using different number of years (blue stars as in figure 3) compared with the payoffs that would have occurred in 2003 (red), 2004(magenta), 2005(green) and the 2003-2005 average payoff (black).

It can be concluded that for the hot July condition (such as the 2003 heat wave), all prices even with the addition of standard deviation, underpriced the final value of the option. More worrying than that, the payoff observed in 2003 exceeded all the VaR values. In 2004 and 2005, the payoffs with standard deviation bounds captured the price. However, the estimations that were based on 51, 30 and 20 most recent years performed the worst. The most interesting case, is when the average of the verification payoffs

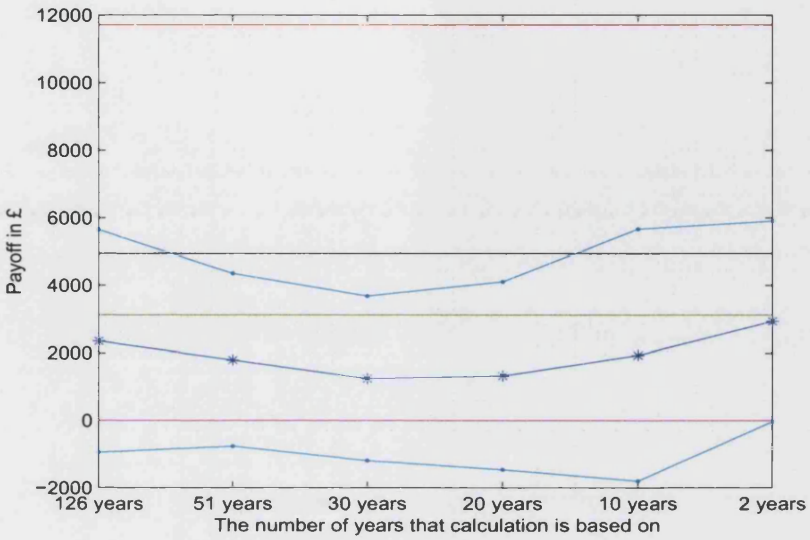


FIGURE 4. Comparing the payoffs of the CDD July Call computed using different numbers of years (blue stars) to the payoffs observed in the verification set (2003 - red, 2004 - magenta, 2005 - green and average between 2003-2005 - black.)

(the average payoff over 2003, 2004 and 2005) was considered - black line on the figure 4. It can be seen that 51, 30 and 20 year based pricing, even with added standard deviation, did not capture the average verification level.

2.2. Comparing characteristics of CDD produced using distribution fitting pricing approach. The next step is to examine what happens to the payoffs, standard deviation and VaR, when Monte Carlo pricing is applied. The Monte Carlo method computes all the measures by using many (in this case 10000) draws from a fitted Normal distribution. Following Zeng's methodology, chapter 2, a Normal distribution is fitted to each historical record, such that the mean and the variance are defined by the

sample mean and variance of each historical data sample. The method is applied using: 126, 51, 30, 20, 10 and 2 years of data.

In figure 5 the payoffs, standard deviation and VaR, obtained using Monte Carlo are compared to payoffs, standard deviation and VaR computed in the previous section using historical data only.

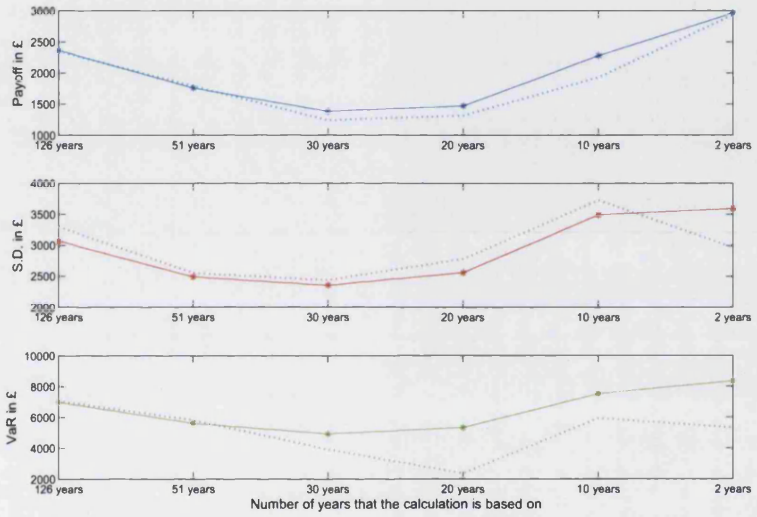


FIGURE 5. Payoffs, standard deviations and VaR measurements for the July CDD Call, computed by fitting a Normal distribution, using CDD mean and variance computed from different number of years. Blue stars are the payoffs in £; red stars are the standard deviation of the payoff in £; finally the green stars are the VaR also in £. We compare the Normal distribution based results to the payoffs (light blue dotted line), standard deviation (pink dotted line) and VaR(dark green dotted line) computed using exclusively historical data.

Years included in calculation	Mean Payout	S.D. of Payout	VaR
126	2328.2	3089.1	7001.70
51	1793.3	2495.4	5649.80
30	1398.90	2341.50	4869.70
20	1466.1	2559.6	5392.30
10	2256	3450.7	7399.20
2	2967.5	3605.1	8326.90

TABLE 2. Payoffs for the CDD July Call computed the Monte Carlo approach by fitting Normal distributions to different numbers of years.

Again, comparing the results for the Berlin option to the the conclusions from the Zeng's paper, it is evident that VaR of the Berlin max temperature CDD option computed using the Monte Carlo method differs the most, when compared to the VaR based on the historical record only. Additionally, payoffs, standard deviation and VaR computed using Monte Carlo exceed the historical record for most yearly samples. This is again consistent with Zeng's result for the US daily average CDD option. Finally, it is also evident that as the number of years that the calculations is based on decreases, the difference between the Monte Carlo based measures and the historical data based measures increases.

So now consider how the Monte Carlo based payoffs compare to the verification payoffs. Again, the comparison is drawn with respect to the actual payoffs for the July 2003, 2004, 2005 and their average payoff.

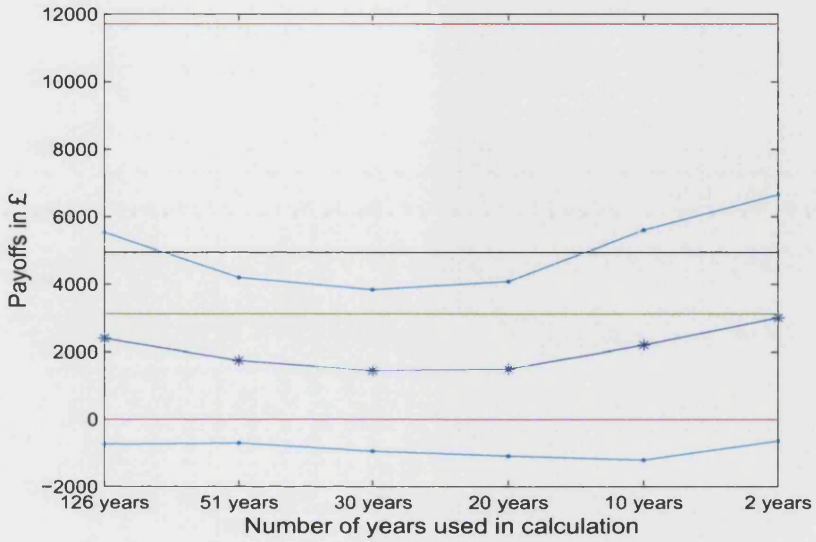


FIGURE 6. Comparing the payoffs of the CDD July Call computed using Normal Distributions fitted to different numbers of years (blue stars) to the payoffs observed in the verification set (2003 - red, 2004 - magenta, 2005 - green and average over all the verification years - black.)

The performance of Monte Carlo derivative prices is very similar to the historical data based prices, as expected. The payoffs computed using Monte-Carlo with 126, 2 and 10 years, capture 2004, 2005 and the average verification payoff within the standard deviation bounds.

3. Pricing the CDD July Call using statistical temperature modelling approaches and ERAP.

In this section Berlin daily maximum temperature based CDD for July, are constructed from synthetic temperature scenarios generated using the methodologies of two papers: M. Davis [**31**] and Cao, Wei [**30**], generated

in chapter 3. For the details on temperature models used in those papers see chapter 3. Then, using the simulated CDD, the average payoff, standard deviations of payoff and VaR are computed in the actuarial pricing framework. In particular the payoff is computed for each scenario and then the average is taken to produce the payoff of the Call option (as for Zeng's pricing). Finally the average payoff, standard deviation and VaR computed based on Davis and Cao and Wei methodology are compared to the average payoff, standard deviation and VaR constructed using the ERAP method (see section 5).

The synthetic temperature scenarios for July, according to the Cao and Wei temperature model were constructed in chapter 3, [**30**]. Here we consider the CDD that resulted from both simulated temperatures constructed using Seasonal Normal Temperature (SNT) and simulated temperatures constructed using Adjusted Seasonal Normal Temperature denoted as $\widehat{Y}_{yr,t}$ (for definitions see section 3). Cao and Wei noted that in order to construct future temperature scenarios SNT should be used, as the Adjusted SNT was specific to certain historical dates. They also suggested, however, that the parameters of the model are more precise when the Adjusted SNT is used. Here we examine how different are the average payoffs, standard deviation of payoffs and VaR for those two approaches.

The synthetic temperature scenarios for July, following the Davis methodology were generated in chapter 3. Again the CDD are constructed from Davis temperature scenarios, by computing the payoff for each scenario and then finding the mean price, standard deviation and the 90th percentile

Method	Mean Payout	S.D. of Payout	VaR
Davis	7353.075158	1636.287952	9475.599341
Cao and Wei SNT	6346.537479	4526.90637	12435.40857
Cao and Wei $\widehat{Y}_{yr,t}$	6692.840868	5726.242814	15155.6974
ERAP	5153.3	5775.99	13575

TABLE 3. Comparing payoffs, s.d. and VaR achieved using Cao and Wei (both SNT and $\widehat{Y}_{yr,t}$ based), Davis and ERAP approach.

(VaR) of those scenario payoffs. The weather options' payoff, s.d. of payoff and VaR are then computed and compared to the ERAP pricing result, figure 7. Table 3 summarises the results obtained for all the methods.

It can be noted that ERAP produces the smallest value of the payoff, compared to the other synthetic temperature simulation methods. At the same time, the ERAP payoff was much closer to the historical data based payoffs (table 1). The measure of standard deviation of scenario payoffs based on ERAP and based on Cao and Wei, were very similar in magnitude to each other. When compared to standard deviation computed using historical data alone (table 1), ERAP and the Cao, Wei method produced much higher values. The levels of VaR also vary between the methods. All the methods produced much higher levels of VaR, when compared to the historical data based VaR.

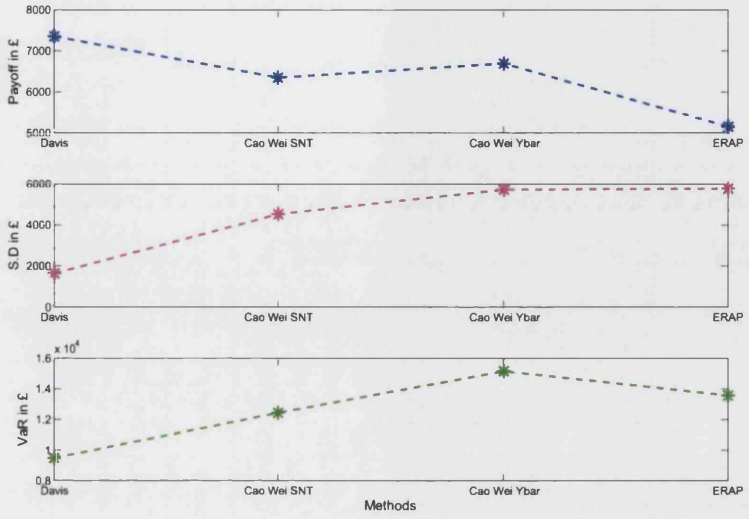


FIGURE 7. Comparing payoffs, s.d. and VaR achieved using Cao and Wei (both SNT and $\widehat{Y}_{yr,t}$ based), Davis and ERAP approach.

4. Conclusion

Now we compare how all the methods performed, when compared to the verification set. It can be seen that the verification payoff that resulted from the hot conditions of 2003 is not captured by any method. In the case of 2004, 2005 and the average of 2003-2005, the ERAP based payoff outperforms other methods.

What we are interested in, is the consistency and reliability of all the pricing methods with respect to the verification set. In order to assess that, differences between the method's average payoffs and the verification's payoffs are summarised in table 4. Additionally, for each method, the area of the payoff plus or minus one standard deviation of that payoff is also considered (table 5).

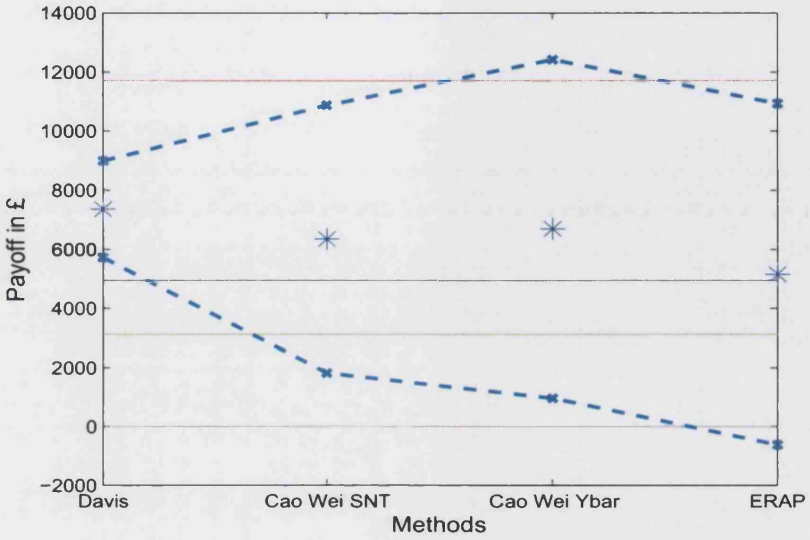


FIGURE 8. Comparison of payoffs constructed using Davis, Cao and Wei and ERAP methodology (blue stars) with the actual payoff of 2003 (red), 2004 (magenta), 2005 (green) and the average of 2003-2005 (black). Cyan lines show the catchment area of standard deviations for each method.

Table 5 is constructed in two stages. First, for each method the lower and the upper boundary is found. The lower boundary is equal to the average payoff minus the standard deviation of the payoff. The upper boundary is computed by adding the payoff and the standard deviation. Then the differences are computed as following:

- if the verification payoff is lower than the lower boundary for the studied method, then the difference is taken between the lower boundary and the verification payoff;

Method	verification - 2003	2004	2005	aver. of 2003-2005
Hist 126	-9355.31	2349.69	-770.31	-2591.98
Hist 51	-9914.04	1790.96	-1329.04	-3150.71
Hist 30	-10462.42	1242.58	-1877.42	-3699.09
Hist 20	-10384.05	1320.95	-1799.05	-3620.71
Hist 10	-9772.27	1932.73	-1187.27	-3008.94
Hist 2	-8770.00	2935.00	-185.00	-2006.67
Davis	-4351.92	7353.08	4233.08	2411.41
Cao and Wei SNT	-5358.46	6346.54	3226.54	6346.54
Cao and Wei $\widehat{Y}_{yr,t}$	-5012.16	6692.84	3572.84	1751.17
ERAP	-6551.70	5153.30	2033.30	211.63

TABLE 4. Differences between the Davis, Cao,Wei and ERAP based payoffs and the verification set.

- if the verification payoff is higher than the upper boundary for the studies method, then the difference is taken between the upper boundary and the verification payoff;
- if the verification payoff is between the lower and the upper boundary, then zero is recorded.

It is clearly evident that ERAP outperforms all the other modelling methods for the Berlin max temperature CDD for July, given the 2003,2004, 2005 and average 2003-2005 verification set, in terms of mean absolute error. It also performed consistently across all 3 verification years. Finally, even in exceptionally hot conditions observed in 2003, the payoff did not cross the

Method	verification - 2003	2004	2005	aver. of 2003-2005
Hist 126	-6056.19	0	0	0
Hist 51	-7358.44	0	0	-595.11
Hist 30	-8024.11	0	0	-1260.78
Hist 20	-7600.20	0	0	-836.87
Hist 10	-6038.55	0	0	0
Hist 2	-5804.54	0	0	0
Davis	-2715.63	5716.78	2596.78	775.12
Cao and Wei SNT	-831.55	1819.63	0	0
Cao and Wei $\widehat{Y}_{yr,t}$	0	966.59	0	0
ERAP	-775.71	0	0	0

TABLE 5. Differences between the s.d. defined lower and the upper boundaries for Davis, Cao,Wei and ERAP and the verification set.

VaR value, which suggests that the ERAP based price should be used instead of the historical data based price, as it produces better risk estimates.

5. Incorporating medium range weather forecast information into pricing and reevaluating weather derivative.

In this last section we briefly discuss, how the methodology studied and tested in chapter 4 should be used, in order to more accurately price weather options. Even if the forecast is not available at the time of pricing a weather option, once the contract starts it is possible to assess the value of it, using the information received from weather forecasts.

In chapter 4 it was shown how to kernel dress and mix ensemble forecasts with the climatology in the perfect model scenario. Medium range weather forecasts (up to 14 days) also come in the form of an ensemble forecast (see section 1.3).

In chapter 5 it has been shown that ERAP produces consistent and, more importantly, enhanced climatology for a time series for both the weather like process and the chaotic Lorenz series. It has also been demonstrated in this chapter, that ERAP based pricing produced more coherent result, when compared to the historical pricing or other statistical models.

This leads to the conclusion that in order to achieve better insight into value of weather options, simulated temperature scenarios should be constructed by mixing kernel dressed physical weather forecast information and the ERAP based temperature simulations, such that the parameters of kernel dressing and the mixing weight between the forecasts and simulated data are determined by minimising Ignorance Skill Score [**45**].

6. Conclusion

We set out on a quest to design a high quality benchmark for long term temperature forecasts beyond the capabilities of a historical climatology. We also wanted to explore how weather risk exposure can be reduced by combining all the available information in terms of weather forecasting including this benchmark. In particular we have chosen a weather derivatives contract as an example of how one can reduce the risk due to weather uncertainty.

In chapter 2 a variety of existing weather forecasting methodologies have been considered together with their strengths and weaknesses. It has been stated that on a short time scale high quality weather forecasts are available. It is also true that the equations of motion chosen by the weather offices to model (approximate) the evolution of the atmosphere exhibite chaotic behavior. As a result, such forecasts contain uncertainty that rapidly grows with time. On a long time scale, historical climatology, statistical temperature simulation and the probability of being above or below some 'norm' were the only options available. Also in chapter 2 a weather risk hedging strategy has been chosen, in the form of a weather derivative. There are several pricing frameworks that can be used in the evaluation of such weather contracts and due to the nature of weather derivatives any pricing framework requires quantitative long term weather forecasts.

In chapter 3 we went deeper into the temperature simulation techniques that have been previously used in evaluating a weather contract. In particular we followed in the footsteps of Cao and Wei [**30**] and Davis [**31**] and applied their methodologies to a chosen data set (Berlin daily temperatures).

The Berlin data was examined in detail where some historical statistic was recorded. Additionally in that chapter a test process was developed that mimicked well the Berlin temperature time series. The purpose here was to build a time series that is non-linear, periodic on multiple time scales, contains random components and at the same time resembles temperature series typically observed in the European climate.

We proceeded by developing (in chapter 4) a methodology that allows to produce combined forecasts. In particular it allows to construct a forecast that contains information form both physical forecasts (forecasts obtained from models of the dynamics of the underlying system that generated the data) and historical/simulated statistics. This methodology has been extensively tested on two noisy chaotic series and a Threshold Autoregressive process. It has been demonstrated that in the case of chaotic systems, in the perfect model scenario with perturbed initial conditions, the quality of the physical forecast (in terms of the amount of information that such forecasts contain about the actual observations) decreases as the forecast horizon increases. This highlighted the importance of the historical/simulated data that is used to compensate for the information loss from physical forecasts.

Given that, for the purposes of weather risk hedging long term weather forecasts are required, our attention has been drawn to the possibilities of improving the quality of long term forecasts by constructing a sophisticated temperature simulation methodology. A methodology that takes into account the dynamics of the weather series and does not assumes that the

'observable' components of the temperature times series (such as 'seasonality', daily fluctuations, trends, and so on) can be defined mathematically and modelled independently. In particular, in chapter 5 an Ensemble Random Analog Prediction (ERAP) climate generator was introduced. The generator was extensively tested both in the perfect model scenario and on the actual Berlin data, the perfect models being a chaotic time series and a weather-like random process, that was specifically designed to mimic real weather data. It was evident that in all those experiments ERAP generated statistics consistent with the statistic of the learning set. In most cases the true statistic of the chosen test process were captured better by the generated ERAP synthetic data than by the limited learning set.

Finally we have applied ERAP for the purpose of pricing weather derivatives and compared ERAP performance, in terms of pricing, to other existing methods for weather index modelling, chapter 6. It has been observed that for all the verification sets the ERAP based pricing performed well, both in terms of mean absolute error and VAR.

6.1. Future Work. This work raises new questions:

- How does the behavior of the information loss in the combined forecasts change with different distributions of the initial perturbations, both in terms of mixing parameter α and the kernel dressing parameter σ?
- How does the estimation of the parameters in the combined forecast change if the optimisation of the parameters is done independently?

- How to best address the parameter estimation for the ERAP technique. Should it be forecast orientated or end application cost function orientated?
- Can ERAP be extended to multidimensional time series? For example temperature and precipitation data combined.
- How can ERAP produced ensembles be transformed into distributions? this could allow more sophisticated optimisation techniques to used for the pricing of weather derivatives.
- How well a global trend is captured by ERAP. This important in the case of the fast Global warming.
- Finally, further investigation is needed to examine weather derivative pricing and weather data simulation using ERAP and combined forecasts under different climates.

Bibliography

[1] *www.metoffice.gov.uk/research/nwp/publications*

[2] *http://www.metoffice.gov.uk*

[3] *Statistical Analysis in Climate Research*, Hans von Stokch and Francis W. Zwiers, Cambridge University Press, Feb 2002

[4] *"Sea Surface temperature for Climate"*, *http* : *//www.metoffice.gov.uk/research/nwp/satellite/infrared/sst/sst$_c$limate.html*

[5] *"The physical Science Basis", Chapter 1: Historical Overview of Climate Change Science*, Intergovernmental Panel on Climate Change (IPCC), Fourth Assessment Report, Working Group I Report.

[6] *"Comparison of Ensemble - MOS Methods using GFS Reforecasts"*, Daniel S. Wilks and Thomas M. Hamill, American Meteorological Society, Jun 2007

[7] *"Prediciting Uncertainty in forecasts of weather and climate"*, T.N. Palmer, ECMWF Technical Memorandum No.294

[8] *Elementary Linear Algebra.* Howard Anton, John Wiley and Sons, 1994

[9] *Companies not forecasting the hidden costs of weather.* Met Office Press, 26th November 2001. The survey have been conducted by Brand and Issue market research company on behalf of the Met Office.

[10] *Statement on Extreme Weather Events*, World Meteorological Organisation, UN, No 695, July 2003

[11] *Climate Extremes: Observations, Modeling, and Impacts*, David R. Easterling et al., Science, Sept 2000, Issue 289, p.2068-2074

[12] *Comment on WMO Statement on Extreme Weather Events,*, Khanderkar, L., EOS Transactions, 2003, AGU, Vol 84, Issue 31, p.428-428

[13] *Weather Report - Climate. A world of weather: Segnificant climatic anomalies and episodic events during 2004.*, Environmental Finance, March 2005, pp.14-15.

[14] Annual industry-wide survey of weather market activity, conducted by the PricewaterhouseCoopers, 2004, on the behalf of the Weather Risk Management Association (WRMA).

[15] *Weather Risk- 5 year in review: Confounding the forecasts*, Mark Nicholls, Environmental Finance, October 2004

[16] *Climate Risk and Weather Market. Financial Risk Management with Weather Hedging.*, Robert S. Dischel, Risk Waters Group Ltd, 2002.

[17] *Financial Calculus. An introduction to derivative pricing*, Martin Baxter and Andrew Rennie, Cambridge University Press, 1996

[18] *Probability with Martingales*, David Williams, Cambridge University Press, 1991

[19] *Introduction to Stochastic Calculus with Applications*, Fima C Klebaner, Imperial College Press, 1998

[20] *Mathematics of Finacial Markets*, Robert J. Elliott and P. Ekkehard Kopp, Springer, 2000

[21] *Stochastic Differential Equations: An Introduction with Applications*, Bernt Øksendal, Springer, 2000

[22] *The pricing of options and corporate liabilities*, F. Black and M. Scholes, Journal of Political Economy 81, 637-654.

[23] *Arbitrage Theory in Continuous Time*, Tomas Björk, Oxford University Press, 1998

[24] *Options, futures and other derivatives*, John C. Hull, Prentice-Hall International, Inc., 1989

[25] *Probability and Statistics*, Morris H. DeGroot, Addison-Wesleu Publishing Company, 1989

[26] *Essentials of Econometrics*, Damodary Gujarati, McGraw-Hill International, 1992

[27] *Chaos and Time Series Analysis*, Julien Clinton Sprott, Oxford University Press, 2003

[28] *The weather generation game: a review of stochastic weather models*, D. S. Wilks and R. L. Wilby, Progress in Physical Geography, Vol. 23, No. 3, 329-357, 1999

[29] *Pricing Weather Derivative*, Lixin Zeng, Journal of Risk Finance, Spring 2000.

[30] *Weather Derivatives Valuation and Market Price of Risk*, M. Cao and J. Wei, The Journal of Futures Markets, Vol 24, No 11, 1065-1089, 2004

[31] *Pricing weather derivatives by marginal value*, Mark Davis, Quantitative Finance Volume 1, 2000, 1-4

[32] *Weather Forecasting for Weather Derivatives*, S.D. Campbell and F. X. Deibold, Journal of the Mathematical Statistical Association, 100, 6-16.

[33] *Caution to the wind*, F. Dornier and M. Queruel, Weather Risk Special Report, Energy and Power Risk Management/Risk Magazine, August 2000

[34] *On Modelling and Pricing Weather Derivatives*, P. Alaton, B. Djehiche, D. Stillberger, Applied Mathematical Finance, Volume 9, Issue 1, 2002, pp. 1-20

[35] *Dynamical pricing of weather derivatives*, D. C. Brody, J. Syroka, M. Zevros, Quantitative Finance Volume 2, Institute of Physics Publishing, 2002, pp. 189-198

[36] *Long-term growth in a short-term market*, E. Farma and J. MacBeth, Journal of Finance 29, 1974, 857-885

[37] *On value preserving and growth optimal portfolios*, R. Korn and M. Schäl, Mathematical Methods of Operations Research 50(2), 1999, 189-218

[38] *The numerair portfolio*, J. Long, Journal of Financial Economics, 1990, 29-69

[39] *Fair Pricing of Weather Derivatives*, E. Platen and J. West, July 2004, working paper.

[40] *Investment Science*, David G. Luenberger, Oxford University Press, 1998

[41] *Optimal Design of Weather Derivatives*, P. Barrieu and N. E. Karoui ALGO Research Quartely, Volume 5, No 1, spring 2002

[42] *Extension of the CPC long-lead temperature and precipitation outlooks to general weather statistics.* Briggs W. M., and D. S. Wilks, 1996, J. Climate, 9, 34963504

[43] *Weather Derivatives and Seasonal Forecast*, S Yoo, 2003

[44] *Journal of Atmospheric Science*, Lorenz E.N. Deterministic nonperiodic flow, **20**, 130-41, 1963

[45] *Scoring Probabilistic Forecasts*, Bröcker J., Smith L.A., Weather and Forecasting 22 (2), 382-388, 2007

[46] *Chaos and Time-Series Analysis*, J.Clinton Sprott, Oxford University Press, New York, 2003

[47] *Numerical Recipies in C. The art of Scientific Computing.*, William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery, Cambridge University Press, 1992.

[48] *Measure, Integral and Probability*, Marek Capinski and Ekkehard Kopp, Springer, 1999

[49] *Probability and Statistics*, Morris H. DeGroot, Addison-Wesley publishing company, 1989

[50] F. Paparella et al Phys.Lett.A (1997), L. A. Smith, Proc International School of Physics "Enrico Fermi", Course CXXXIII, page 177-246, Italian Physical Society (1997)

[51] *Weather Derivative Valuation - The Meteorological, Statistical, Finacial and Mathematical Foundations*, S. Jewson, Anders Brix, with C. Ziehmann, Cambridge University press, 2005.

[52] *Nonlinear Time Series Analysis*, Holger Kantz and Thomas Schreiber, Cambridge University Press, 1997

[53] *Singular vector ensemble forecasting system and the prediction of flow dependant unceratinty*, S. Jewson, M. Ambaum, C. Ziehmann, 2004

[54] *Non-linear Time Series Anlaysis. A Dynamical System Approach*, Howell Tong, Clarendon Press Oxford, 2003.

[55] *"Probabilistic Forecasts and their verification."*, Zoltan Toth, Environmental Modelling Centre, NOAA/NWS/NCEP, Ackn: Y. Zhu, O. Talagrand, http://wwwt.emc.ncep.noaa.gov/gmb/ens/index.html

[56] Matlab's functions specification website: www.mathworks.co.uk