EXECUTIVE SUMMARIES

The Executive Summary consists of a brief description of the problem followed by a largely equation-free summary of the progress made and the results obtained by the study group.

VISUAL ANIMAL

Industry representative:
Terence van Zyl, University of the Witwatersrand, Johannesburg

Moderator:
Matthew Woolway, University of the Witwatersrand, Johannesburg

Student moderators:
John Atherfold and Krupa Prag, University of the Witwatersrand

Executive Summary

Designing an efficient method to track and count game/livestock by uniquely identifying individuals in a nature conservationist setting, was the problem considered by the study group. Many animals may be individually recognised by their bio-metric characteristics in the same way as human fingerprints uniquely identify human beings. Human fingerprints are not able to be used for mutual visual recognition as this physical bio-metric feature is relatively small. However, most animal skin patterns are large enough physical bio-metric features which can be used for mutual visual recognition without technical aids - at least theoretically.

The precursory stages of approaching the problem included ingesting, cleaning and exploring the input data set, a collection of 1000 animal images. Undesirable images were removed from the data set; these included corrupted, blurred and/or images which contained multiple animals. The remaining images were than resized to equal dimensions and were used in designing a method to uniquely identify each individual.

Image Processing (IP) techniques were applied as preliminary means to uniquely identify the animal in the images. Initially Region Agency Graphs were used, which locally cluster pixels or regions in an image and uniquely label them. Regions from the edge of the segmented image were selected and used to identify the background color which was to be removed from the images with the intend to solely retain the animal in the image. However, due to the similarity of the background colours to the animal, this method failed to provide unique individual identification. Recognising that individuals can be uniquely identified by their stripes, a 'barcode extraction' method was setup. This method entailed horizontally scanning the centre of the

image and using colour intensity to extract only the dark stripes from the image and transforming them to a frequency domain to represent a unique 'barcode'. From the obtained results, it was found that this method has potential; however, due to massive human requirement on the tweaking of parameters, selecting cropping location, and accounting for scale and rotational invariance, this was not seen as a tractable approach.

The images considered from the data set have non-uniform backgrounds, animals positioned at different angles, and animals which camouflaged with their backgrounds. Due to these characteristics of the images, IP techniques are challenging to successfully apply. Thus, machine learning algorithms were considered, particularly Transfer Learning was utilised in the study. The weightings from a pre-trained neural network from the Serengeti project were used in the Siamese neural network which was developed. The Siamese network uses Similarity Learning. Images are paired up, where the paired images are either of the same animal, or different animals. These pairs are labeled and given to the pre-trained network for training. The metric used to tune the model parameters were a function of the euclidean distance between the pair of images, and whether the pair of images were of the same animal. Only the final layer of the neural network was tuned during training. Preliminary results indicate that the Male-Female classification has $\approx 90\%$ out-of-sample accuracy for the verification of the network using Transfer Learning, and $\approx 70\%$ error-rate in sample accuracy for the verification of the Siamese network using Similarity Learning.

The conclusion is that machine learning techniques produced better results in comparison to the applied IP techniques. The designed machine learning model can be improved by refining the hyper-parameter choices by using optimisation techniques, obtaining larger training data sets, applying multiple loss functions, as well as further prepossessing the data into age and/or sex classes to possibly improve classification schemes.

OPTIMIZING URBAN FARMING

Industry representative:
Norman Hoeltz, aparate

Moderator:
Montaz Ali, University of the Witwatersrand, Johannesburg

Student moderator:
Matthews Sejseo, University of the Witwatersrand, Johannesburg

Study group members:
Jeandre Boshoff, F. De Villiers, Adriana Roux, Azwindini Maphiri, Michael Olusanya, Evans Thulare, Tendani Mutavhatsindi, Shaun Sepuru, Bill Seota and Clarinda Nhangumbe

Executive Summary

Around the world urban farming has significantly increased. Johannesburg is no exception to this phenomenon. Food gardens are springing up on the inner city roof-tops. The motivation is due to basic food security, employment creation and a multitude of socio-economic opportunities. Unlike traditional farming, urban farming yields small crop sizes and is spread throughout the city. It can be done year round as it is farmed under controlled conditions. The produce of farms has to be collected and transported to the market. There is a large number of small farms in contrast to the norm of one or two large farms. The challenges that arise in this system are: no access to information and no coordinated logistics amongst the farms.

Urban farmers spend considerable time in coordinating logistic and transporting their produce to the market. A sizeable amount of their produce is destroyed before it reaches its destination. In business terms this translates to a significant loss of revenue. There is a need of a support tool that can advise on the type of crops to plant, and also determine the optimal planting and harvesting time based on the price of produce in the market. Once the harvest is done, a coordinated system amongst farms has the potential to improve the logistical operations and transportation of produce to the market.

The Study Group modelled the problem in two parts. Firstly, the predictive pricing model was considered. The model is intended to help the farmers to decide what crops to grow, when to grow them and when to harvest and sell the produce.

Multiple products and historical price data was provided by the industry representative. The group proceeded by first mining and organizing the data, followed by analyzing the data to obtain some insight about price behavior of different products in the market. The price of four products - cabbage, potatoes, carrots and strawberries where singled out for the analysis. The prices were shown to be fluctuating, with no distinctive patterns, and there were no co-relations between prices of different crops. The analysis of data was followed by building the predictive model. The model based on ARIMA and machine learning was proposed. The results of the model still need to be improved.

Secondly, coordinated logistics and transportation of produce to market. The objective here was to develop an optimization model that can be used for route planning and scheduling for collection of produce from the farms. The model has to take into account various constraints such as certain produce require a refrigerated vehicle, the vehicle must start and end at the deport, the limit on the maximum distance each vehicle can travel. The problem was first solved using the heuristic method. We cluster the farms into three groups, since three vehicles are available. A vehicle was dedicated to serve a specific group. The local search algorithm 2-opt was used to find the optimal route given a specific cluster. This approach gave a suboptimal solution. To improve on this a mathematical model was developed. The problem was formulated as a split-multi commodity vehicle routing problem. Split - farms are allowed to be visited more than once; multi commodity we are dealing with multiple produce. Work still has to be done to improve the above results, where the main points are:

(1) formulate and find an effective solution of the predictive price model for planting and harvesting times,

(2) find an efficient solution of the mathematical model for route planning,

(3) possibly combine the two models to have a better coordinated logistics system among the farms.

ROGUE WAVES

Industry representative:
Thama Duba, Durban University of Technology, Durban

Moderators:
Colin Please, University of Oxford, United Kingdom and Graeme Hocking, Murdoch University, Perth, Australia

Student moderator:
Kendall Born, University of the Witwatersrand

Study group member:
Meghan Kennealy

Executive Summary

For safe passage of ships, it is necessary and vital to avoid weather conditions where the ship may be put in danger. A serious problem is that there are "Rogue Waves", whose amplitude is unusually large for a given sea state that damage or even sink ships. Reports suggest that these waves appear and disappear suddenly and unexpectedly. Trying to give information that can identify situations where these rouge waves might occur would be very helpful in assisting how to plan shipping routes to ensure safety. The Study Group was asked to explore mechanisms that might create rouge waves and to consider if some general rules could be found that might explain where and when rogue waves might occur.

There is no precise definition of a rogue wave but a common rule is that such a wave is larger than the general surrounding wave field. Much of the modelling to predict rogue waves considers long waves in deep water. The study group mainly focused on the nonlinear Schrödinger Equation (NLSE) to model rogue waves. Considerable analysis exists for the NLSE. The Study Group pursued the route considered by Cousins. The main idea is motivated by the knowledge that there is a very special solution to the NLSE, namely a soliton. This represents a set of waves whose amplitude is only significant in a particular region and this pulse travels at a constant speed and unchanging shape or height.

The Study Group analysed a NLSE and was able to find a range of values for which the solution would produce large amplitudes within a reasonable time frame. The

group also found values that produced large waves over a long period of time, and values that produced short waves over a small time period. To understand the full mechanisms that generate rogue waves and to predict them was not possible in the duration of the Study Group due to the complexity of the problem. The group is hopeful that enough work was completed such that in following years the problem can be continued further.

SUGAR CRYSTAL ANALYSIS

Industry representative:
Richard Loubser, Sugar Milling Research Institute, University of KwaZulu-Natal, Durban

Moderator:
Bill Seota, University of the Witwatersrand, Johannesburg

Study group members:
Emmanuel Netshiunda, God'sgift Uzor, Risuna Nkolele, Precious Makganoto, Andrew Paskaramoorthy, David Merand and Lucky Nouralden

Executive Summary

Introduction

Sugar crystal (SC) analysis - SC counting and dimension measurements of sugar crystals plays an essential role in the quality control stage of commercial sugar production. The final product needs to be of reliable quality for the end consumer; hence the raw sugar must meet industry-standard requirements. More specifically, drawing insights about the distribution of the size of the crystals, their shape and the density (number of crystals per millilitre) is required in optimising of the growth of sugar crystals in the production of fine sugar. The size and shape of crystals are also used to control the various production processes and to determine the presence of impurities. The Growth Rate Dispersion (GRD) is a term that describes the range of growth rate exhibited by sugar crystals. High GRD values become problematic since the growth rate of a crystal determines the eventual size of a crystal. The GRD determines the distribution of the size of the final sugar produced. The Study Group found that the measurement of sugar crystal sizes, their distributions and their density can be efficiently achieved using crystal image-processing techniques. We attempt to develop an automated pipeline for processing images of sugar crystals and subsequently counting the number of crystals in an image.

Problem definition

Conventional methods of SC analysis involve human inspection and the use of sieves. The time and labour invested into this activity make it a costly task, while the results produced by it are of limited quality. Analysis concerning the shape of crystals involves determining key parameters, including the elongation ratio, the coefficient variation and the mean aperture. Additionally, the density provides important information about the refinement process. There could be more than a few ways in which to frame the problem, one of which could be through understanding the mechanics of the fluids within which the crystals are immersed. While that may be a feasible approach to the solution, it is one that could require many steps before completing the analysis. An intuitive approach takes the form of an image processing point of view, with the aim of automating[1] the analysis of SC analysis from images taken by a microscope (pictomicrographs).

Methodology

We needed several key components that constitute a fluid pipeline. Our process implemented the following steps:

1. Conversion of the image to Gray Scale
2. Filtering by using Image Blurring
3. Segmentation by using Thresholding
4. Object detection
5. Object classification

Pre-processing

Pre-processing involves getting the images to a state where the embedded objects are easy to manipulate. For the most part, the steps that follow aim to achieve an image that is noise free: We converted each image to gray scale, applied average filtering to reduce noise, used contrast stretching, and finally, dilated the image's pixels to remove the residual noise.

Segmentation

Segmentation refers to partitioning regions of interest in the image, where each region contains an object. An important part of this procedure is separating the background from the (perceived) foreground objects. Common segmentation trans-

[1]Not necessarily an official term

formations include edge detection, the watershed transformation, and thresholding procedures. Each approach was attempted, but Otsu's method of thresholding (with binarisation) demonstrated the strongest in-sample performance amongst alternative methods.

Object detection

Following segmentation, the remaining objects (of which most were crystals) were identified for the purpose of counting and assessing shapes. Contours were drawn around foreground objects to count crystals.

Crystal detection

The aim of crystal detection was to prevent impurities and other visual anomalies from being perceived as crystals by the algorithm. The detection part of the algorithm places markers on the boundaries of objects of interest (objects perceived to be crystals, owing to the algorithm's specification).

While this pipeline of algorithms does not produce error-free results, the onus lies with factory operators to weigh the trade-off between using methods such as the one proposed (or variations thereof) and the alternative manual methods.

MODELLING THE SURFACE WATER POLLUTION IN LAKE KIVU

Industry representative:
Denis Ndanguza, University of Rwanda, Rwanda

Moderator:
Neville Fowkes, University of Western Australia, Perth, Australia

Student moderators: Nomsa Ledwaba, University of Cape Town and Anastacia Dlamini, University of South Africa

Executive Summary

Lake Kivu, on the border of Rwanda and the Democratic Republic of Congo, is known to be a dangerous lake primarily because of the vast quantities of carbon dioxide and methane dissolved in the deep waters of the lake which, if spontaneously released, could cause massive damage and loss of life. However there are other hazards posed by the levels of pollutants in the surface layers of the lake caused primarily by activities around the lake. Some, having their source in the water itself, include municipal sewage and wastewater treatment plants, toxic runoff from farms and factories, either near the lake or upstream. Other sources are atmospheric and include gas vented into the atmosphere during harvest time and smokestacks from factories near the lake. The released gases and particles eventually end up in the lake either directly or after rain and can effect the health of inhabitants, animals and water life. The aim of the project was to model surface water pollution in the lake.

The group felt no direct progress could be made on this very worthy problem without specifics; what pollutant, what source, and then data/information concerning the primary factors affecting dispersal need to be known at least crudely such as winds, topography and climate. With the above in mind the group decided to set up a generic model that may be used to crudely describe the atmospheric dispersal from a factory chimney. The simple models developed could then be combined with absorption models to determine how much of the released gas/particles would then accumulate and be dispersed within the lake, or flow through and out of the lake. Such lake dispersal models were not developed at the MISG but can be simply modelled. All such models can provide crude estimates for the effect of a known

source on the environment and are an appropriate start for any investigation of any pollution investigation. More detailed monitoring/computational models then need to be developed but of course these are expensive (mainly because the data input needed is expensive to collect) and require major and continuous computational monitoring and processing. In general terms the crude generic models are sensible initially to see if there is a problem and then if there is a problem then a more thorough/detailed/computationally intensive/model is required.

Simple models will be presented in the report and their limitations indicated.

DOUBLE-DIFFUSIVE CONVECTION IN LAKE KIVU

Industry Representative:
Denis Ndanguza, University of Rwanda, Rwanda

Moderator:
David Mason, University of the Witwatersrand, Johannesburg

Student Moderators:
Osman Noreldin, University of KwaZulu-Natal, Pietermaritzburg and Thomas Roy, University of Oxford, United Kingdom

Study Group Members:
Anastacia Dlamini, Gideon Fareo, Neville Fowkes, Masood Khalique, Nomsa Ledwaba, Thato Malema, Azwindini. Maphiri and Bothwell Marefere

Executive Summary

Lake Kivu is situated on the border between Rwanda and the Republic of Congo. It is a stratified lake with four layers of slightly different density. The lake contains dissolved methane and carbon dioxide. The major part of the dissolved gas is in the lowest layer kept there by the weight of the water above. Methane is extracted from the lowest layer to generate electricity.

In Lake Kivu the temperature first decreases with depth for about the first 70 m because of the heating of the lake surface by the sun. Due to geothermal heating from below, the temperature then steadily increases from about 23°K to about 26°K at the bottom of the lake. This produces a negative temperature gradient and describes a fluid heated from below which is destabilising. Dissolved carbon dioxide increases the density of water while dissolved methane decreases the density. Salt (impurities) in the lake increase the water density. The net effect is that there is a steady increase in the water density with depth. This negative density gradient is stabilising because the heavier fluid is below. When the diffusivity of heat is greater than the diffusivity of salt double-diffusive convection can occur. In double-diffusive convection, if a parcel of fluid is displaced upwards it will lose heat at a greater rate than salt and buoyancy forces will drive it back towards its original position. This produces an oscillatory motion in which the direction of the buoyancy force is reversed in each half-cycle with the parcel of fluid overshooting its equilibrium position.

The Study Group was asked to initiate a study of double-diffusive convection in Lake Kivu and investigate the opposing effects of thermal diffusion and mass diffusion.

The Study Group first reviewed the linear stability analysis of a fluid with opposing gradients of temperature and density [1]. The dispersion relation is in the form of a cubic equation. The derivation of the regions of stability and instability in the (Rs, Ra)-plane was reviewed where Ra and Rs are the thermal and salinity Rayleigh numbers. The overstability region in which the instability sets in as an oscillation of growing amplitude was identified. In this region Ra and Rs are both positive, the temperature gradient is unstable and the salinity gradient is stable. Another region of instability is the salt finger region in which both Ra and Rs are negative. The density gradient is unstable and the temperature gradient is stable. The stability of nonlinear double-diffusive convection has been analysed by Huppert and Moore [2].

The Study Group also reviewed the part played by double-diffusive convection in the formation of temperature steps which have been measured experimentally in Lake Kivu. Instead of smooth temperature and salinity gradients, well mixed layers separated by sharp interfaces are found. No temperature steps have been observed above 120 m depth but between 120 m and 460 m depth, the temperature profile has 250 to 350 well mixed layers. The thickness of most layers is between 0.2 m and 1.0 m with interfaces between 0.05 m and 0.4 m and temperature steps between 0.001°K and 0.005°K [3]. In the model of Turner [4] a layer with stable salt gradient is heated from below. A lower convecting layer is first formed which at time t has thickness $t^{1/2}$. The heat and salt in this layer is well mixed by convective heating. There is a diffusive thermal boundary layer ahead of the advancing front of the convecting layer. The thermal boundary layer is assumed to become unstable in an overstable oscillation. The instability criterion in terms of Ra and Rs is given by linear stability theory. The characteristic distance in Ra and Rs is the thickness of the thermal boundary layer. At the onset of instability the lower layer stops growing, which determines the thickness of the bottom layer, convection begins and the second layer starts growing. This can be extended to the formation of many layers [5].

The microstructure profiles of Lake Kivu can lead to an understanding of the macrostructure of the lake. Since temperature steps are formed in the lake at 120 m to 460 m depth it indicates that the instability sets in as an overstability which determines the region of the Rs, Ra plane in which the lake lies at these depths. The large temperature step at 280 m to 300 m of over 1°K may be as a result of the merger of several double-diffusive layers to form one large layer [6]. The heat and salt fluxes through the layers formed by double-diffusive convection could be used to estimate these fluxes through the lake as a whole which is important in estimating the upward flux of heat, salt and gases through the lake [3]. An effective thermal conductivity for the whole series of layers in the temperature steps can be defined which is directly proportional to the thermal conductivity of the lake [5].

References

[1] Bains PG and Gill AE. On thermohaline convection with linear gradients. J Fluid Mech 1969, **37**, 289-306.

[2] Huppert H and Moore DR. Nonlinear double-diffusive convection. J Fluid Mech 1976, **78**, 821-854.

[3] Schmid M, Halbwachs M, Wehrli B and Wüest A. Weak mixing in Lake Kivu: New insights indicate increasing risk of uncontrolled gas eruption. Geochemistry, Geophysics, Geosystems 2005, Volume 6, Number 7. doi:10.1029/2004GC000892.

[4] Turner JS. The behaivour of a stable salinity gradient heated from below. J Fluid Mech 1968, **33**, 183-200.

[5] Huppert H and Linden PF. On heating a stable salinity gradient from below. J Fluid Mech 1979, **95**. 431-464.

[6] Schmid M, Tietze K, Halbwachs M, Lorke A, McGinnes D and Wüest A. How hazardous is the gas accumulation in Lake Kivu? Acta Vulcanologica 2003, **15**, 115-122.

MODELLING CARBON CAPTURE

Industry representative:
Tim Meyers, Centre de Recerca Matematica, Barcelona, Spain

Moderator:
Tim Meyers, Centre de Recerca Matematica, Barcelona, Spain

Student moderator:

Study group members:
Chelsea Bright, Megan Fowlds, Thato Malema, Quinton Ndlovu, Marcia Moremedi and Saul Hurwitz

Executive Summary

The working group began by searching the literature for methods on carbon capture and the background to the problem. It was decided to study an experimental set-up where gas is slowly passed through a narrow column containing an adsorbent material. Mathematical models for this situation are studied in a number of papers and reviews. The governing equations described the uptake of carbon dioxide and nitrogen via a concentration equation with a sink term which was coupled to a temperature equation. The strength of the sink term comes from experimental data. Suitable parameter values were found from the various papers.

The system was non-dimensionalised leading to the group's first, most simple model. This neglected the temperature variation and so resulted in two first order PDEs for the concentration and variation of sink strength. Despite its simplicity a numerical solution was required.

A numerical solution including temperature variation was also investigated. Preliminary results showed qualitatively similar behaviour to experiments. Initially all carbon is removed: the amount passing through the column then slowly increases until the adsorbent is saturated and no further removal occurs. The nitrogen behaviour shows the remarkable feature that over 120% of the initial value exits

the column. This is explained by the fact that at entry the gas consists of CO_2 and nitrogen, at exit it is almost all nitrogen. Hence the proportion at exit is much greater than at entry.

Future work should concentrate on verifying the governing equations (doubts have been raised), seeking numerical solutions and using these to guide the design of carbon capture devices.

DESIGN OF FISHING EXCLUSION ZONES

Industry representative:
Philip Broadbridge, La Trobe University, Melbourne, Australia

Moderator:
Ashleigh Hutchinson, University of the Witwatersrand, Johannesburg

Student moderator:
Michael M^{c}Phail, Oxford University, United Kingdom

Study group members:
Colin Please, Ebrahim Fredericks, Bothwell Maregere and Roy Gusinow

Executive Summary

With international recognition that natural marine resources are under pressure from over-exploitation, Marine Protection Areas, most of which exclude fishing, are becoming more widespread. A Marine Protected Area (MPA) is an aquatic zone where fishing and other forms of human intervention are regulated and restricted. A No-Take Area (NTA) is created within an MPA, whereby no fishing is permitted, allowing populations a safe region to reproduce and grow. The purpose of these exclusion zones is to enable fish populations to recover from over-fishing and repopulate quickly so that fishing activities can be sustained. Fishing communities have hindered the development and expansion of these exclusion zones. Support can be gathered from these communities by optimizing these zones as well as showing increased fish populations in neighbouring, unregulated regions of NTAs.

The aim of the Study Group is to develop a mathematical model that ensures the survival of endangered fish by achieving balance between fishing, birth, and movement rates. Existing models are improved upon and predictions are made. The reaction-diffusion equation is used to model a single fish species. The influence of the geometry of the exclusion zone is investigated for simple circular and rectangular geometries. A linear stability analysis is conducted in order to obtain the conditions under which the fish population is sustained. By analysing fish behaviour, a Gaussian distribution is used to model the source term that represents fish reproduction. A 1-D model that considers the spacing between exclusion zones is obtained. Existing data is used to determine unknown parameters in the model so that accurate predictions can be acquired.

From the report the following conclusions were made: a circular geometry is the optimal shape for the survival of the fish; a series of exclusion zones is more effective in maintaining the fish population as opposed to one larger exclusion zone; and, in order to adapt the model to make it more realistic, further analysis on the data describing the fish behaviour is required.