2025美国数学建模竞赛培训

华南理工大学数学学院
刘深泉教授

目录

- 美国大学生数学建模竞赛题目浏览
- 2020 ICM Problem D：Teaming Strategies
- 2022 MCM Problem C：Trading Strategies
- 2010 MCM Problem A: The Sweet Spot
- 2022 ICME: Forestry for Carbon Sequestration
- 2025 Summary: MCM and ICM

2022 MCM/ICM美国大学生数学建模竞赛

- A题：自行车骑手功率分配建模Power Profile of a Cyclist
- B题：水与水电共享Water and Hydroelectric Power Sharing
- C题：交易策略Trading Strategies
- D题：数据分析系统成熟度评估Data Paralysis? Use Our Analysis!
- E题：林业的碳封存Forestry for Carbon Sequestration
- F题：国际公平及小行星采矿政策研究All for One and One (Space) for All!

2023 MCM/ICM美国大学生数学建模竞赛

- A题：饱经旱灾的植物群落Drought-Stricken Plant Communities
- B题：重新构想马赛马拉 Reimagining Maasai Mara
- C题：预测Wordle结果Predicting Wordle Results
- D题：优先考虑联合国可持续发展目标Prioritizing the UN Sustainability Goals
- E题：光污染Light Pollution
- F题：绿色国内生产总值Green GDP

2024 MCM/ICM美国大学生数学建模竞赛

- A题：资源可用性和性别比例 Resource Availability and Sex Ratios
- B题：寻找潜水器 Searching for Submersibles
- C题：网球运动中的动力 Momentum in Tennis
- D题：五大湖水资源问题 Great Lakes Water Problem
- E题：财产保险的可持续性 Sustainability of Property Insurance
- F题：较少非法野生动物的贸易 Reducing Illegal Wildlife Trade

2020 ICM Problem D：Teaming Strategies

This data set was processed from a much larger dataset covering nearly 2000 matches
from five European national soccer competitions, as well as the 2018 World Cup

As societies become more interconnected, the set of challenges they face have become increasingly complex. We rely on interdisciplinary teams of people with diverse expertise and varied perspectives to address many of the most challenging problems. Our conceptual understanding of team success has advanced significantly over the past 50+ years allowing for better scientific, creative, or physical teams to address these complex issues. Researchers have reported on best strategies for assembling teams, optimal interactions among teammates, and ideal leadership styles. Strong teams across all sectors and domains are able to perform complex tasks unattainable through either individual efforts or a sequence of additive contributions of teammates.

One of the most informative settings to explore team processes is in competitive team sports. Team sports must conform to strict rules that may include, but are not limited to, the number of players, their roles, allowable contact between players, their location and movement, points earned, and consequences of violations. Team success is much more than the sum of the abilities of individual players. Rather, it is based on many other factors that involve how well the teammates play together. Such factors may include whether the team has a diversity of skills (one person may be fast, while another is precise), how well the team balances between individual versus collective performance (star players may help leverage the skills of all their teammates), and the team’ s ability to effectively coordinate over time (as one player steals the ball from an opponent, another player is poised for offense).

To respond to the Huskie coach′ s requests, your team from ICM should use the provided data to address the following:

In light of your modeling skills, the coach of the Huskies, your home soccer (known in Europe and other places as football) team, has asked your company, Intrepid Champion Modeling (ICM), to help understand the team′ s dynamics. In particular, the coach has asked you to explore how the complex interactions among the players on the field impacts their success. The goal is not only to examine the interactions that lead directly to a score, but to explore team dynamics throughout the game and over the entire season, to help identify specific strategies that can improve teamwork next season. The coach has asked ICM to quantify and formalize the structural and dynamical features that have been successful (and unsuccessful) for the team.

The Huskies have provided data[1] detailing information from last season, including all 38 games they played against their 19 opponents (they played each opposing team twice). Overall, the data covers 23,429 passes between 366 players (30 Huskies players, and 336 players from opposing teams), and 59,271 game events.

Create a network for the ball passing between players, where each player is a node and each pass constitutes a link between players. Use your passing network to identify network patterns, such as dyadic and triadic configurations and team formations. Also consider other structural indicators and network properties across the games. You should explore multiple scales such as, but not limited to, micro (pairwise) to macro (all players) when looking at interactions, and time such as short (minute-to-minute) to long (entire game or entire season).

Identify performance indicators that reflect successful teamwork (in addition to points or wins) such as diversity in the types of plays, coordination among players or distribution of contributions. You also may consider other team level processes, such as adaptability, flexibility, tempo, or flow. It may be important to clarify whether strategies are universally effective or dependent on opponents′ counter-strategies. Use the performance indicators and team level processes that you have identified to create a model that captures structural, configurational, and dynamical aspects of teamwork.

Use the insights gained from your teamwork model to inform the coach about what kinds of structural strategies have been effective for the Huskies. Advise the coach on what changes the network analysis indicates that they should make next season to improve team success.

Your analysis of the Huskies has allowed you to consider group dynamics in a controlled setting of a team sport. Understanding the complex set of factors that make some groups perform better than others is critical for how societies develop and innovate. As our societies increasingly solve problems involving teams, can you generalize your findings to say something about how to design more effective teams? What other aspects of teamwork would need to be captured to develop generalized models of team performance?

The ball passing network is an undirected graph.

$$
\begin{array}{c c}
 & \begin{array}{cccc} 1 & 2 & \cdots & 11 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ \vdots \\ 11 \end{array} &
\begin{pmatrix}
w_{11} & w_{12} & \cdots & w_{1,11} \\
w_{21} & w_{21} & \cdots & w_{2,11} \\
\vdots & \vdots & \cdots & \vdots \\
w_{11,1} & w_{11,2} & \cdots & w_{11,11}
\end{pmatrix},
\end{array}
\tag{1}
$$

In this paper, we assume that the ball passing network is an *undirected graph*, which means the edge has no direction. Therefore, we have the number of passes between players $w_{ij} = w_{ji}$

Meanwhile, the ball passing network can be vividly illustrated in **Figure 1** by *Gephi*.

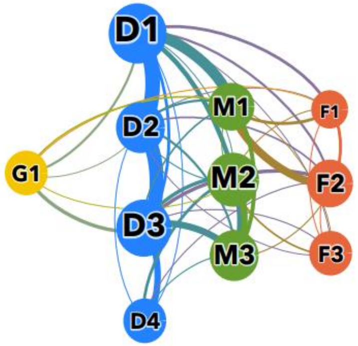
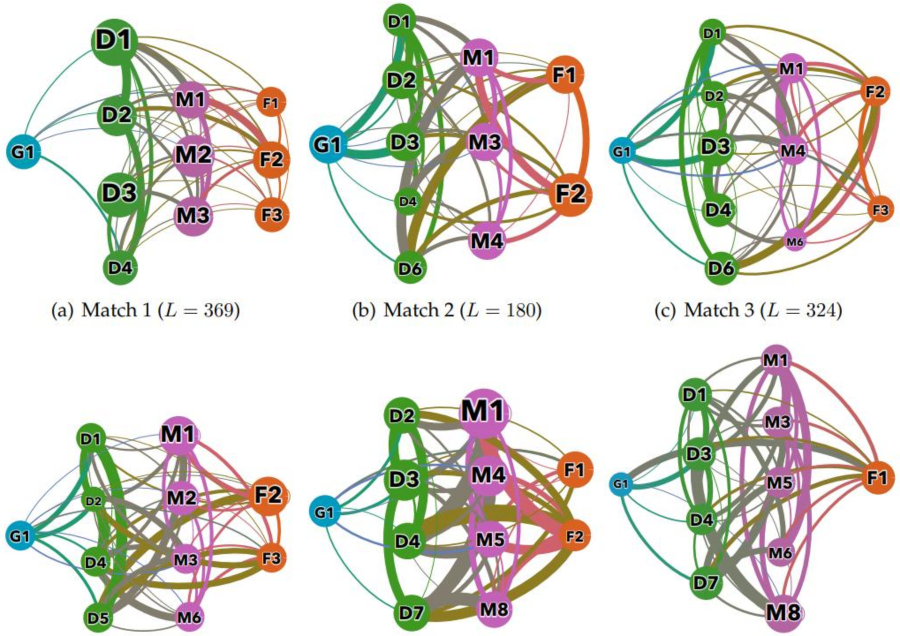


Figure 1: An example of the ball passing network



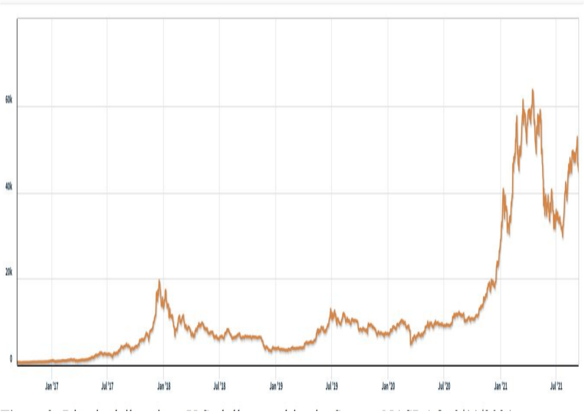
(a) Match 1 ($L = 369$)
(b) Match 2 ($L = 180$)
(c) Match 3 ($L = 324$)
(d) Match 4 ($L = 354$)
(e) Match 10 ($L = 359$)
(f) Match 13 ($L = 347$)

Figure 5: Entire Game and Entire Season Network Passes

2022 MCM Problem C：Trading Strategies

Background Market traders buy and sell volatile assets frequently, with a goal to maximize their total return.There is usually a commission for each purchase and sale. Two such assets are gold and bitcoin.
Figure 1: Gold daily prices, U.S. dollars per troy ounce. Source: London Bullion Market Association, 9/11/2021
Figure 2: Bitcoin daily prices, U.S. dollars per bitcoin. Source: NASDAQ, 9/11/2021

Attachments
THE TWO DATA FILES PROVIDED CONTAIN THE ONLY DATA YOU SHOULD USE FOR THIS PROBLEM. LBMA-GOLD.csv BCHAIN-MKPRU.csv
Data Descriptions
LBMA-GOLD.csv Date: The date in mm-dd-yyyy (month-day-year) format.
USD (PM): The closing price of a troy ounce of gold in U.S. dollars on the indicated date.BCHAIN-MKPRU.csv
Date: The date in mm-dd-yyyy (month-day-year) format.
Value: The price in U.S. dollars of a single bitcoin on the indicated date.



2022 MCM Problem C：Trading Strategies

Requirement
You have been asked by a trader to develop a model that uses only the past stream of daily prices to date to determine each day if the trader should buy, hold, or sell their assets in their portfolio. You will start with $1000 on 9/11/2016. You will use the five-year trading period, from 9/11/2016 to 9/10/2021. On each trading day, the trader will have a portfolio consisting of cash, gold, and bitcoin [C, G, B] in U.S. dollars, troy ounces, and bitcoins, respectively. The initial state is [1000, 0, 0]. The commission for each transaction (purchase or sale) costs α% of the mount traded. Assume α g o l d α_{gold}α gold = 1% and α b i t c o i n _{bitcoin}α
bitcoin = 2%. There is no cost to hold an asset.
Note that bitcoin can be traded every day, but gold is only traded on days the market is open, as reflected in the pricing data files LBMA-GOLD.csv and BCHAIN-MKPRU.csv. Your model hould account for this trading schedule. To develop your model, you may only use the data in the two spreadsheets provided: LBMA-GOLD.csv and BCHAIN-MKPRU.csv.
Develop a model that gives the best daily trading strategy based only on price data up to that day. How much is the initial $1000 investment worth on 9/10/2021 using your model and strategy? Present evidence that your model provides the best strategy. Determine how sensitive the strategy is to transaction costs. How do transaction costs affect the strategy and results? Communicate your strategy, model, and results to the trader in a memorandum of at most two pages. Your PDF solution of no more than 25 total pages should include: ne-page Summary Sheet. Table of Contents. Your complete solution. One- to two-page Memorandum. Reference List.
Note: The MCM has a 25-page limit. All aspects of your submission count toward the 25-page imit (Summary Sheet, Table of Contents, Reference List, and any Appendices). You must cite the sources for your ideas, images, and any other materials used in your report.

The Price Prediction Model: time series forecasting model

ARIMA:Autoregressive Integrated Moving Average model

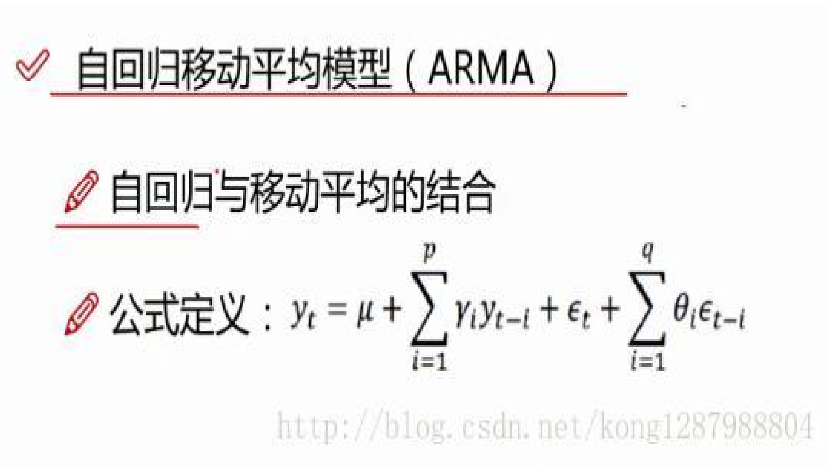
自回归(AR)模型;

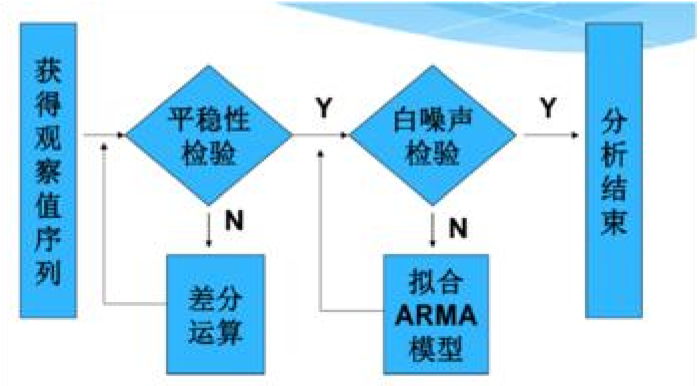
移动平均(MA)模型;

自回归移动平均(ARMA)模型;

自回归整合移动平均模型(ARIMA);

季节性整合自回归移动平均模型(SARIMA)模型。





Forecasting - Linear Model： Autoregressive Integrated Moving Average

The ARIMA (p,d,q) model is known as the Autoregressive Integrated Moving Average Model, where p is the autoregressive term; d is the number of differences when the time series is stationary; q is the number of moving average items. This model is a combination of autoregressive (AR) and moving average (MA), which can transform a non-stationary time series into a stationary time series, and then regress the lagged values of the dependent variable, the present and lagged values of the random error term to the model established. The formula is given in equation 2.

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{2}$$

Where y_t is the current value; μ is the constant term; γ_i is the autocorrelation coefficient; ε_t is the error term; θ_i is the coefficient of error term. The ARIMA (p,d,q) model process for furcating asset price is shown in Figure 3.

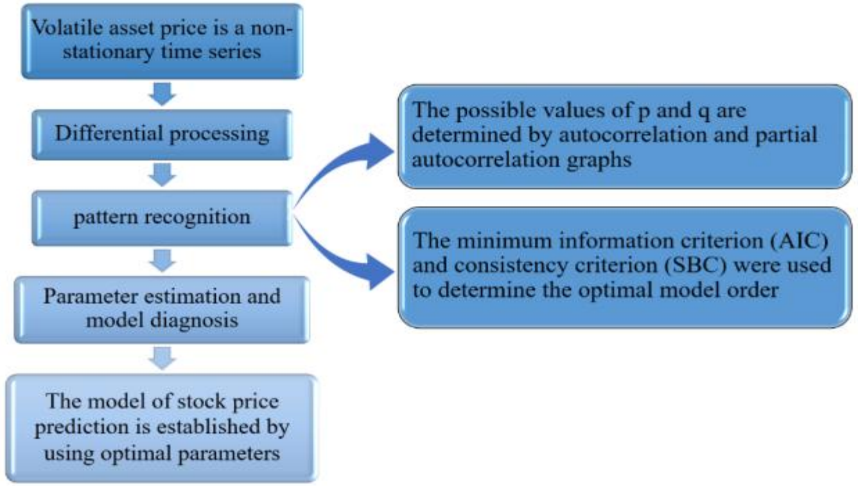


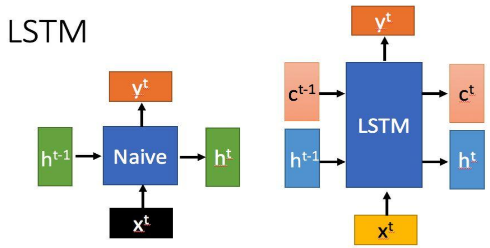
Figure 3: ARIMA Forecast Asset Price Flow Chart

The Price Prediction Model: LSTM - Long short-term memory

LSTM : Long short-term memory

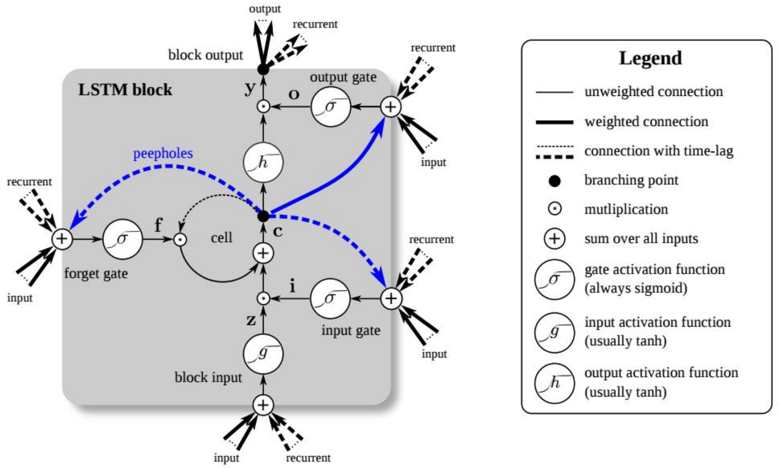
A Long short-term memory (LSTM) is a type of Recurrent Neural Network specially designed to prevent the neural network output for a given input from either decaying or exploding as it cycles through the feedback loops. The feedback loops are what allow recurrent networks to be better at pattern recognition than other neural networks. Memory of past input is critical for solving sequence learning tasks and Long short-term memory networks provide better performance compared to other RNN architectures by alleviating what is called the vanishing gradient problem.

LSTMs due to their ability to learn long term dependencies are applicable to a number of sequence learning problems including language modeling and translation, acoustic modeling of speech, speech synthesis,speech recognition, audio and video data analysis, handwriting recognition and generation, sequence prediction, and protein secondary structure prediction.



c change slowly ➡ c^t is c^{t-1} added by something

h change faster ➡ h^t and h^{t-1} can be very different



Forecasting NonLinear Model: LSTM Neural Network Nonlinear Prediction Model

Long Short Term Memory Network (LSTM), a modified recurrent neural network, can handle the problem of long-range dependencies.

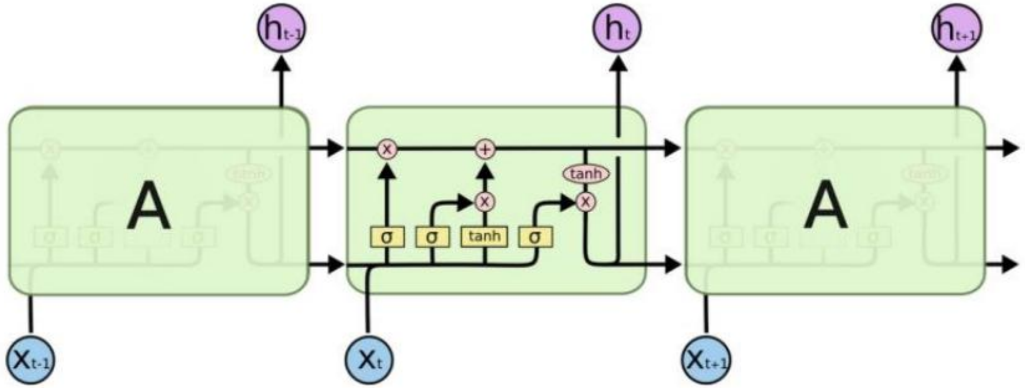


Figure 4: LSTM Working Mechanism Flow Chart

The key to the LSTM is the rectangular box in the second elliptical rectangle in Figure 4, which is called the memory block and contains three main gates (forget gate, input gate, output gate) and a memory cell. The horizontal line at the top of the box is called cell state, which is like a conveyor belt that controls the transfer of information to the next moment. The two tanh layers in the diagram above correspond to the input and output of the cell.

2010 MCM A: The Sweet Spot

- Explain the “sweet spot” on a baseball bat.
- Every hitter knows that there is a spot on the fat part of a baseball bat where maximum power is transferred to the ball when hit. Why isn’t this spot at the end of the bat? A simple explanation based on torque might seem to identify the end of the bat as the sweet spot, but this is known to be empirically incorrect. Develop a model that helps explain this empirical finding.
- Some players believe that “corking” a bat (hollowing out a cylinder in the head of the bat and filling it with cork or rubber, then replacing a wood cap) enhances the “sweet spot” effect. Augment your model to confirm or deny this effect. Does this explain why Major League Baseball prohibits “corking”?
- Does the material out of which the bat is constructed matter? That is, does this model predict different behavior for wood (usually ash) or metal (usually aluminum) bats? Is this why Major League Baseball prohibits metal bats?

2010 MCM A: The Sweet Spot Model

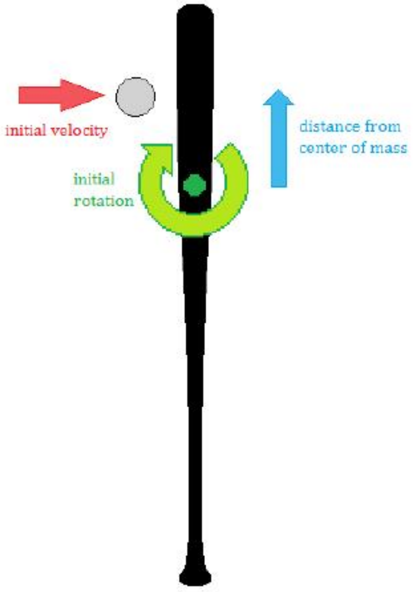


Figure 1. The collision.

The ball collides with the bat at a distance l from the center of mass of the bat. We assume that the collision is head-on and view the event such that all the y-component velocities are zero at the moment of the collision. After the collision, the ball has a final velocity v_f and the bat has a final linear velocity V_f and an angular velocity ω_f at the center of mass.

When the ball hits the bat, the ball briefly compresses and decompresses, converting kinetic energy to potential energy and back. However, some energy is lost in the process, that is, the collision is inelastic. The ratio of the relative speeds of the bat and the ball before and after the collision is known as the *coefficient of restitution*, customarily designated by e: $e = 0$ represents a perfectly inelastic collision, and $e = 1$ means a perfectly elastic one. In this basic model, we make two simplifying assumptions:

Our equations are a discretized version of the Euler-Bernoulli equations:

$$\rho \frac{\partial^2 y(z,t)}{\partial t^2} = F(z,t) + \frac{\partial^2}{\partial z^2}\left(YI \frac{\partial^2 y(z,t)}{\partial z^2} \right),$$

where

ρ is the mass density,

$y(z,t)$ is the displacement,

$F(z,t)$ is the external force (in our case, applied by the ball),

Y is the Young's modulus of the material (a constant), and

I is the second moment of area ($\pi R^4/4$ for a solid disc).

2022 ICM E: Forestry for Carbon Sequestration

- Background

Climate change presents a massive threat to life as we know it. To mitigate the effects of climate change, we need to take drastic action to reduce the amount of greenhouse gases in the atmosphere. Simply reducing greenhouse gas emissions is not enough. We need to make efforts to enhance our stocks of carbon dioxide sequestered out of the atmosphere by the biosphere or by mechanical means. This process is called carbon sequestration. The biosphere sequesters carbon dioxide in plants (especially large plants like trees), soils, and water environments. Thus, forests are integral to any climate change mitigation effort.

Forests sequester carbon dioxide in living plants and in the products created from their trees including furniture, lumber, plywood, paper, and other wood products. These forest products sequester carbon dioxide for their lifespan. Some products have a short lifespan, while others have a lifespan that may exceed that of the trees from which they are produced. The carbon sequestered in some forest products combined with the carbon sequestered because of the regrowth of younger forests has the potential to allow for more carbon sequestration over time when compared to the carbon sequestration benefits of not cutting forests at all.

At the global level, forest management strategies that include appropriate harvesting can be beneficial for carbon sequestration. However, overharvesting can limit carbon sequestration.Forest managers must find a balance between the value of forest products derived from harvesting and the value of allowing the forest to continue growing and sequestering carbon as living trees. In doing so, they must consider many factors such as age and types of trees,geography, topography, and benefits and lifespan of forest products.

The concerns of forest managers are not limited to carbon sequestration and forest products. They must make forest management decisions based on the many ways their forest is valued.These may include, but are not limited to, potential carbon sequestration, conservation and biodiversity aspects, recreational uses, and cultural considerations.

2022 ICM E: Forestry for Carbon Sequestration

- Requirements
- The International Carbon Management (ICM) Collaboration has been formed to develop guidance for forest managers around the world trying to figure out how to utilize and manage their forests. One-size-fits-all guidance is simply not possible as the make-up of forests, climates, populations, interests, and values vary widely around the world.
- Develop a carbon sequestration model to determine how much carbon dioxide a forest and its products can be expected to sequester over time. Your model should determine what forest management plan is most effective at sequestering carbon dioxide.
- The forest management plan that is best for carbon sequestration is not necessarily the one that is best for society given the other ways that forests are valued. Develop a
- decision model to inform forest managers of the best use of a forest. Your model should determine a forest management plan that balances the various ways that forests are
- valued (including carbon sequestration).
- To better understand your model, consider some of the following questions, as well as questions of your own:

* What is the spectrum of management plans that your decision model may suggest?

* Are there any conditions which would result in a forest that should be left uncut?

* Are there transition points between management plans that apply to all forests?

* How are characteristics about a specific forest and its location used to determine transition points between management plans?

2022 ICM E: Forestry for Carbon Sequestration

- Apply your models to various forests. Identify a forest that your decision model would suggest the inclusion of harvesting in its management plan.

• How much carbon dioxide will this forest and its products sequester over 100 years?

• What forest management plan should be used for this forest? Why is this the best approach?

• Suppose the best management plan includes a time between harvests that is 10 years longer than current practices in the forest. Discuss a strategy for transitioning from the existing timeline to the new timeline in a way that is sensitive to the needs of forest managers and all who use the forest.

• Some people believe we should never cut down any trees and yet you identified a forest that should include harvesting in its management plan. Write a one- to two-page non_x0002_technical newspaper article explaining why your analysis identified including harvesting in the management of this forest rather than it being left untouched. Ultimately, your article should convince the local community that this is the best decision for their forest.

The modeling of Forestry for Carbon Sequestration

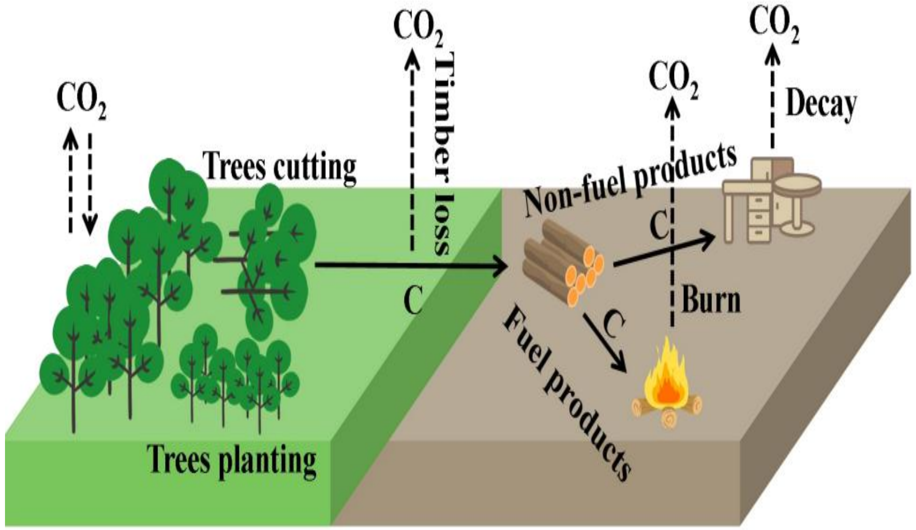
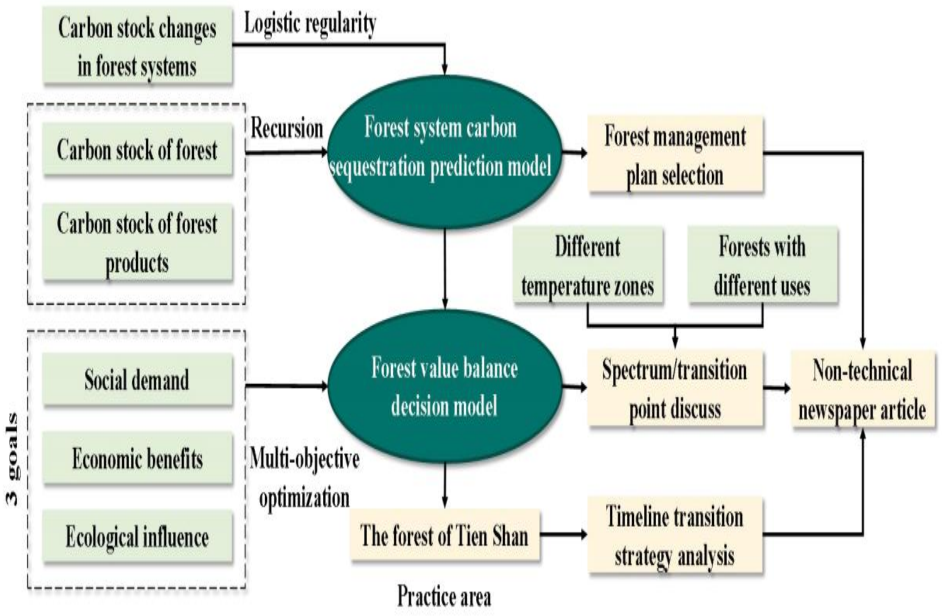


Fig.2 Schematic diagram of carbon cycle of forest system (forest and forest products)



MCM and ICM Summary

- MCM:The Mathematical Contest in Modeling
- Problem A: 连续型建模问题
- Problem B: 离散型建模问题
- Problem C: 数据处理型建模问题
- ICM:The Interdisciplinary Contest in Modeling
- Problem D: 运筹学或网络科学问题
- Problem E: 环境科学类问题
- Problem F: 政策类问题
- Reading and Practice the MCM-ICM Model