PROJECT 2

Submit your project through *WISEflow*. Please do not share nor reproduce questions or answers of this project in other internet websites or digital platforms. The submission deadline is Tuesday October 24th, at 14:00 hrs. The project can be done individually or in a group of at most 2 students. No cooperation between people who are not submitting this project as a group is allowed. It is possible to change groups throughout the semester and it is also possible to do some project(s) alone and other project(s) in a group. Provide **all your AMPL files** (model code, data, running commands, solution file, etc.) compressed in a single file (.zip). Include all files needed to run all parts of the project, even if from one to another task the changes are just marginal (we need all files to be able to run without modifying what you submitted). In addition, provide a written report with your model formulations and the answers to the questions required in each part. The formulation of your models can be typed in a text editor (e.g. Word, LaTeX), written by hand and scanned, or copied directly as text or screenshot from the AMPL code files when it applies (please just be careful the presentation must be clear enough for a reader). In the written report, it is fine that when there is just a marginal change from one task to another, in the latter you include just the modified part of the formulation (e.g., in task 2 you just defined a new variable or modified one constraint of the model you formulated in task 1, then it is fine that you included the full model formulation in task 1 and only the new variable definition and new constraint that you modified in task 2). Provide a short description (no more than two sentences, e.g. "#demand fulfillment") for every objective function and constraint in your formulations. All model formulations in this project, either involving continuous and/or integer variables, must be **linear**. Only as a reference (not as a requirement), the expected length of your report is: Part A half of a page, Part B half of a page, Part C three pages, Part D four pages.

Part A (10%)

The major airline *AirBAN* aims at delivering services to passengers in the BAN402 network of 18 countries that we designed in Lecture 1. Currently, an important decision is to decide which countries in the network will serve as hubs. Aware of your expertise in decision modelling, they ask you for help to make such important decision. In particular, they believe you have studied carefully the model formulations of the 11 problems we studied in lectures 7 and 8, namely: fixed cost, knapsack with a single resource, knapsack with multiple resources, facility location as a mixed integer linear programming model, facility location as a pure integer programming model (0/1-problem), assignment, set partitioning, set covering, set packing, symmetric TSP, and asymmetric TSP.

Let I be the set of 18 countries. Establishing a hub in country i implies a cost c_i, information given to you as data. *AirBAN* is considering different approaches to decide which country (or countries) will serve as hub(s). These approaches are described in the two tasks below (each task is independent from each other).

1. A first approach is to define a configuration of hubs such that the total cost is minimized, while securing that every country in the network can be reached from at least one hub in a flight time of up to $\bar{t}$. The flight time between country i and country j is equal to $t_{i,j}$, defined for all $18 \times 18 = 324$ possible pairs (i,j) in the network. The values of $\bar{t}$ and $t_{i,j}$ are given to you as data. Which one of the 11 model formulations would you select to support *AirBAN* in this approach? How would you use the given data to set the values of the parameters in this formulation? Could this model be infeasible?

2. In another approach, *AirBAN* wants to consider not only the cost of establishing the hubs, but also some administrative costs associated to manage the network. The structure of these administrative costs is as follows. There are 18 local offices, one in each country of the network. Each local office is led by a local manager. In addition, every hub will have a hub manager. Each local manager will have to report to exactly one hub manager in the network. If a hub in country i is selected as the hub to which the local manager of country j must report, *AirBAN* incurs a cost equal to $r_{i,j}$. This cost parameter is defined for all $18 \times 18 = 324$ possible pairs (i,j) in the network. Due to managerial limitations, if a hub is established in country i, a maximum of m_i local managers will be allowed to report to this hub. *AirBAN* gives you data with all the values of $r_{i,j}$ and m_i, and asks you to find a configuration of hubs,

such that the cost of establishing hubs and administrative costs in total would be minimized. Which one of the 11 model formulations would you select to support *AirBAN* in this approach? How would you use the given data to set the values of the parameters in this formulation? Could this model be infeasible?

Part B (20%)

In the article "Developing Optimal Student Plans of Study" (available on a link in the *Complementary readings* folder in *Canvas*), the author describes the use of an integer linear programming model to help students plan their studies at an educational institution. The model is explained verbally on pages 415-416 of the article, and the mathematical formulation of the model is outlined in the Appendix.

Suppose there are some courses that are well-known to be hard. Let us define the set *CHard* containing all those hard courses (of course, *CHard* is a subset of set C). The willingness to take these courses and whether to combine several of them or not in the same term vary across students. We focus here on three students: Nils, Lisa, and Abel. As they have heard you are an expert in decision modelling, they have asked you to modify the model to help them design their plans of study. Your formulations must be linear and may involve new definitions (e.g. of variables), new expressions (e.g. in the objective function and/or constraints), the modification of some expressions in the original formulation, etc. These must be formulated in mathematical terms (not in AMPL code). The new situations are described below (each situation is independent from each other).

1. In Nils' opinion, the best strategy is to take at most one hard course per term. Which modification(s) would you introduce in the model to help Nils?
2. Lisa has no problem to take two or more hard courses in the same term. However, she would like to make sure that such situation does not happen in more than two terms in total during her studies. Which modification(s) would you introduce in the model to help Lisa?
3. Abel likes terms with exactly two hard courses. He tells you please to add a reward of 10 units in the objective function per each term in which exactly two hard courses appear in his plan of study. Which modification(s) would you introduce in the model to help Abel?

Part C (30%)

The company *Oil&BAN* refines two types of crude oil, which after a sophisticated process result in saleable products. The process is summarized as follows.

The crude oil is converted to components in the Crude Distilling Units (CDUs) located in the refining department of the company. According to the characteristics of the CDUs and of the types of crude oil, the refining process will provide proportions of different components. This is illustrated with a simple numerical example in Figure 1. Here, one unit of crude oil 1 will separate into 60% of component 1 and 40% of component 2 if it goes through CDU 1. The corresponding proportions will be 70% and 30% respectively in CDU 2. Each crude oil will give different proportions of components.

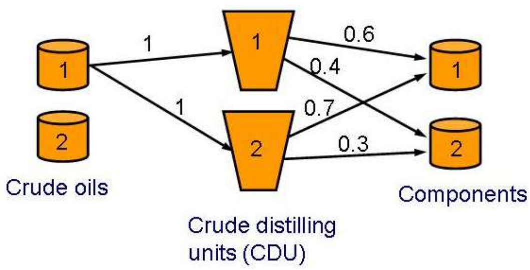

Figure 1: A numerical example of the CDU process.

In general, we will refer by $A_{i,b,j}$ to the amount of component b obtained from refining one unit of crude oil j in CDU i. The amount of crude oil j arriving each day, which we will denote by E_j, is fixed according to the agreements of the company with its suppliers. The crude oil is ready to be used at the CDUs within the same day of arrival or, alternatively, it can be stored at the refining department.

The company has two CDUs at its refining department. If a CDU operates on a given day, it must refine at least *minR* and at most *maxR* units of crude oil on that day. The cost of refining one unit of crude oil j in CDU i is $Cref_{j,i}$. Alternatively, the company may prefer to not operate a CDU during one day. However, in order to keep daily production at the company, everyday there must be at least one CDU operating.

The CDUs are subject to maintenance routines periodically. In our problem, we will consider a planning time horizon of 12 days. There must be at least one day of maintenance for each CDU during the last six days of this planning horizon. The cost of a maintenance routine at a CDU is *CMain*. If a CDU is subject to maintenance routine on a given day, it cannot process crude oil during that day.

All the components generated are sent to the blending department of the company, except for one component called *lowqfuel*. Assume that the components sent from the refining department on day t will arrive at the blending department at the beginning of day $t+1$. The cost of transporting one unit of any component from the refining to the blending department is $Ctra1$. It is not possible to store components at the refining department, so the total amount of all components generated at the refining department are sent to the blending department, except for *lowqfuel* which is sent directly from the refining department to depots. This is because *lowqfuel* is already a saleable product, although at relatively low price in the markets. The cost of transporting one unit of *lowqfuel* from the refining department to depot d is $Ctra2_d$. Assume that the *lowqfuel* sent from the refining department on day t will arrive at the depots at the beginning of day $t+1$.

When the components arrive at the blending department, they can be stored in tanks or mixed according to recipes for generating final products. The recipe for producing one unit of product p needs $R_{b,p}$ units of component b. The cost of producing one unit of product p is $Cprod_p$.

From the blending department, the products are sent to depots (there is no storage of products at the blending department). The cost of transporting one unit of any product from the blending department to depot d is $Ctra3_d$. Assume that what is produced in the blending department on day t will arrive at the depots at the beginning of day $t+1$.

Once the saleable products arrive at the depots (including *lowqfuel*), they are ready to be shipped to the markets. The cost of shipping one unit of any product from depot d to market k is $Ctra4_{d,k}$. Alternatively, the products may be stored at the depots.

There is a maximum demand limit for product p from market k in day t, which we will refer by $\delta_{p,k,t}$. Assume that what is shipped from the depots on day t will arrive at the markets at the beginning of day $t+1$ (thus, in order to satisfy daily demand the company needs to ship one day in advance). Also, assume that it is possible to partly fulfil demand of a same market by supplying from different depots. The price of one unit of saleable product p in all markets is S_p.

The cost of storing one unit of any type of crude oil at the refining department is $Cinvj$ per day. The cost of storing one unit of any type of component at the blending department is $Cinvb$ per day. The cost of storing one unit of any type of product at depot d is $Cinvp_d$ per day. (Note all these inventory costs are incurred per any unit stored at the end of each day.)

Figure 2 illustrates the different stages of this supply chain.

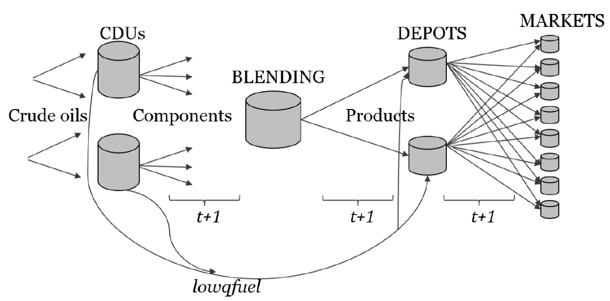

Figure 2: Illustration of the supply chain of the company *Oil&BAN*.

1. Formulate a mixed integer linear programming model for the multi-period planning problem of the company, including decisions on refining, maintenance, production, transportation, storage and sales. The objective is to maximize the total profit over the planning horizon. Implement the model in AMPL and solve it using the solver ***cplex*** and the data instance contained in the file "DataC.dat" (you may modify the file according to your own definitions). Note the set of time periods (expressed in days) is defined as $T = \{0, 1, 2, \ldots, 12\}$, although the relevant decisions are from periods 1 to 12. The initial

inventory of saleable products at the depots (that is, at the end of period $t = 0$), is given in the two columns under the header $Izero_{p,d}$ in Table 1. Assign value zero to any other initial inventory or initial flow variable that you may require in your implementation. However, the company would like to have at least some minimum quantity of components and products on inventory at the end of the planning horizon, in order to anticipate future demand. The final inventory of all components must be at least 100 units each, except *lowqfuel* for which there is no minimum requirement. The minimum inventory of products for each depot at the end of the planning horizon is given in Table 1 under the header $Ifinal_{p,d}$.

	$Izero_{p,d}$		$Ifinal_{p,d}$	
	D1	D2	D1	D2
premium	200	220	225	225
regular	420	450	325	350
distilF	70	120	150	150
super	130	220	250	200
lowqfuel	0	0	0	0

Table 1: Initial inventory and minimum final inventory quantities of each saleable product at each depot.

a) How much is the optimal profit? When do the CDUs operate and when do they not operate? When do the maintenance routines take place?
b) Is there unsatisfied demand for some saleable product(s)? If so, briefly explain why this occurs.
c) Is there any inventory of crude oils left at the end of the planning horizon? If so, briefly explain why this occurs.
d) Is for some component(s) the inventory left at the end of the planning horizon above 100? If so, briefly explain why this occurs.

2. Suppose there are some daily fixed costs associated to the operation of the CDUs. The fixed cost structure is such that if only CDU1 operates on a given day, the fixed cost that day is 60000; if only CDU2 operates on a given day, the fixed cost that day is 120000; if both CDU1 and CDU2 operate on a given day, the fixed cost that day is 135000. Modify your model formulation in order to represent this new situation (keeping linearity). Implement the new model in AMPL and solve it using the solver ***cplex***. How much is the optimal profit? When do the CDUs operate and when do they not operate? When do the maintenance routines take place?

Note: When solving mathematical programming models, it is interesting to observe how the algorithms of the solvers approach the optimal solution during the optimization process and some statistics on the dimension of the problem and the solution time. For this purpose, when using *cplex*, you can add in your .run file the following lines anywhere before the statement *solve*:
option solver cplex;
option cplex_options 'mipdisplay=4';
option show_stats 1;

Part D (40%)

A number of mathematical programming models have been used in the context of humanitarian logistics and disaster response. In fact, whenever a natural disaster occurs, public authorities are in need of deploying operations to assist inhabitants in the affected regions. For the interested readers, the article "Humanitarian logistics and emergencies management: New perspectives to a sociotechnical problem and its optimization approach management" (available on a link in the *Complementary readings* folder in *Canvas*) presents a comprehensive overview on the related literature.

The introduction above motivates us to study the distribution of relief items in the Momiss region, which has just been affected by a big earthquake. Figure 3 shows the map of Momiss. The sheet "Traveling" of the file "DataD.xlsx" contains data on the travel times for a representative network of Momiss, which includes the 16 *cities* surrounded either by a circle or a square in the map of Figure 3. Notice the matrix of data is asymmetric, which indicates that the time to travel from one to another city in the network might in some cases not coincide with the time to travel between these two cities in the opposite direction. Our task is to plan the distribution of relief items within this network of 16 cities, considering some conditions. Firstly, the relief items will be handled in four operation centres within the network. For this purpose, a previous study by the public authorities has indicated that only seven of the 16 cities can act as operation centres. These are surrounded by a square in Figure 3, namely: N1, N3, C1, C2, S2, S4, and S5. Among these candidate locations, our solution must decide which four cities are selected to serve as operation centres (at most one operation centre may be installed per city). Also, there are four trucks available to perform the distribution. The trucks will initially be placed in the four different cities that act as operation centre (one truck per centre). A certain amount of time is needed to set up and load a truck before a trip starts, which slightly differs among the candidate cities to serve as operation centres. These loading times are given in the sheet "Loading" of the file "DataD.xlsx". We need to make sure that every city is visited exactly once by some of the truck. Due to capacity limitations, every truck must visit exactly four cities (including the one it departs from). Every truck must get back to its city of origin after having completed the distribution.

1. Formulate a mixed integer linear programming model to simultaneously decide which cities should be selected as operation centres and which routes the trucks should traverse within the network, so that the longest time needed by a truck is minimized.[1] Implement and solve the model in AMPL, using the solver ***gurobi***. Which cities are selected as operation centres? How much is the longest time needed by a truck in your solution? From which operation centre does (do) this (these) truck(s) depart? In your written report, outline the optimal routes that you found, either graphically (e.g. drawing it in the figure), as a table, commented in words, or any other way that could easily be understood by an external person. (Imagine you are communicating the solution to the public authorities from Momiss, who are not necessarily familiar with mathematical programming, so they want to clearly understand the final solution instead of seeing a raw display with the optimal value of your variables.)

2. As an alternative metric to the longest time, suppose that the public authorities have constructed a list of impact categories, taking into account how strong has the earthquake hit the different cities of Momiss. In this list, the cities have been partitioned into three categories. The first category corresponds to cities suffering a *high-impact*, which include the five southern cities surrounded by a red circle or a red square in Figure 3, namely: S1, S2, S3, S4, and S5. The second category corresponds to cities suffering a *medium-impact*, which include the six central cities surrounded by a blue circle or a blue square in Figure 3, namely: C1, C2, C3, C4, C5, and C6. The remaining cities, surrounded by a green circle or a green square in Figure 3, are categorized as *low-impact*. Based on these impact categories, the authorities have computed an *emergency score* to represent the benefits of visiting a city early in the routes of the trucks. These scores are as follows:

   - If a high-impact city is the initial point of a truck (i.e., the city is selected as operation centre), it contributes with a score of 200; if a high-impact city is the second city in the trip of a truck, it contributes with a score of 160; and if a high-impact city is the third city in the trip of a truck, it contributes with a score of 120.
   - If a medium-impact city is the initial point of a truck (i.e., the city is selected as operation centre), it contributes with a score of 150; if a medium-impact city is the second city in the trip of a truck, it contributes with a score of 100; and if a medium-impact city is the third city in the trip of a truck, it contributes with a score of 50.

   An *overall emergency score* is computed as the sum of the score contributions included in the distribution plan. Note that the low-impact cities are not considered in the calculation of the overall score. Also, note that the last city visited by a truck (before coming back to its city of origin) does not contribute to the overall score, no matter the impact category of the city.[2]
   Modify your model for this new situation, as to maximize the overall emergency score (keeping linearity).

[1] Note the time needed by a truck includes the loading time and the time it requires to complete its route. Also, note that the task focuses on the longest time among all four trucks. For example, a solution where three trucks use 20 hours and one truck uses 100 hours is worse than a solution where four trucks need 80 hours, because the longest time in the first solution is 100 hours and the longest time in the second solution is 80 hours (and 80 is less than 100).

[2] For example: if the first truck starts in S5, and then goes to S2, and then to C3, and then to N2, and then goes back to S5, it contributes with 200 + 160 + 50 + 0 = 410 to the overall impact score; if the second truck starts in C1, and then goes to C4, and then to N5, and then to S4, and then goes back to C1, it contributes with 150 + 100 + 0 + 0 = 250 to the overall impact score. These contributions add up to 410 + 250 = 660, and we would still need to add the contributions of the two remaining trucks to compute the overall emergency score in this example.

Implement and solve the model in AMPL, using the solver ***gurobi***. What is the overall emergency score in your solution? Which cities are selected as operation centres? Outline the optimal trips that you found (in the same way as you did in task 1).

3. What is the overall emergency score of the solution that you found in task 1? Since this solution was found without considering the data on the impact of the earthquake, it perhaps does not perform well in the overall emergency score metric. Attempting to capture both the longest time and the overall emergency score criteria, you propose to incorporate a new condition to the model of task 1, which assures that the overall emergency score is at least an $\alpha\%$ of the optimal overall emergency score found in task 2. For example, if the optimal overall emergency score found in task 2 was 1000, and you use $\alpha = 90\%$, the overall emergency score in the solution found by this new model should be at least 900. How do you incorporate this condition in the model (still keeping linearity)? Consider the following six different values of α: 100%, 95%, 90%, 85%, 80%, 75%. Implement and solve the new model in AMPL for these six data instances. In your written report, summarize your results in a table with three columns showing for each of the six instances the values of alpha, the overall emergency score, and the longest time needed by a truck.

Note: Similarly as we noted in Part C for *cplex*, when using *gurobi*, you can add in your .run file the following lines anywhere before the statement *solve*:
option solver gurobi;
option gurobi_options 'outlev=1';
option show_stats 1;

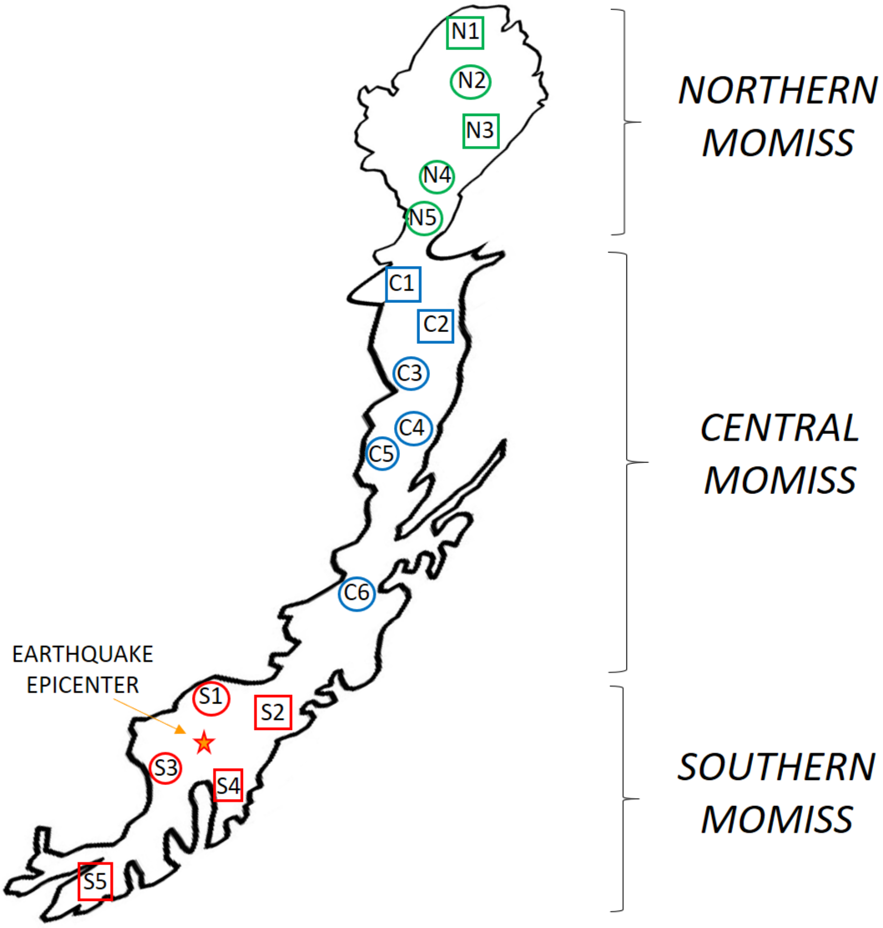

Figure 3: Geographical distribution of the places in Momiss under study.