INATAS TECHNOLOGY AND DATA ANALYSIS SIX CASES

1. SOFT-SENSING – WASTE WATER TREATMENT
2. CAPACITY PLANNING AND DEMAND FORECASTING
3. MANUFACTURING QUALITY AND SYSTEM CONTROL
4. MAINTENANCE AND FAULT DETECTION
5. DEVELOPMENT OF A MEDICAL DIAGNOSTIC TOOL
6. BUS SERVICES AND UTILISATION OF ASSETS

THE PURPOSE OF THESE SLIDES

To demonstrate the benefits involved with the use of Inatas' technology and proprietary software we have outlined 6 cases where the technology has been or is on its way to be used for solving real life problems.

Inatas' software is based on generalizations of Bayesian network theory. Its sophisticated system modelling, predication and decision support/system automation processes put to use large data and expert knowledge to solve difficult analytical problems.

inatas System Modeling Data Analysis Predictive Analytics Automated Decision Making Process Optimization

CASE 1

Detection of phosphate in waste water

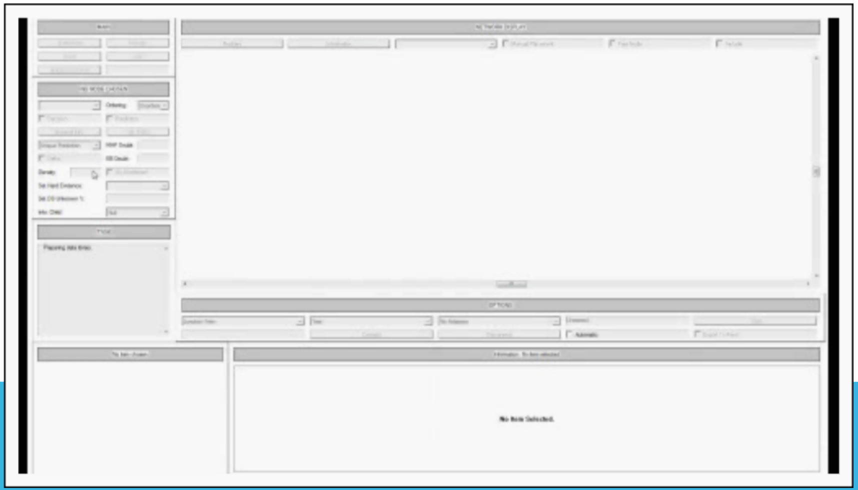
PROBLEM TO BE SOLVED

Tracking the level of phosphate in waste water treatment plants is essential and expensive.

It is desirable to be able to estimate the level of phosphate from observations of cheaper, easily observable variables.

This requires the development of a system model (the relation of the chosen variables to the level of phosphate), a dynamics model (how the system evolves over time) and observation model (the relation between the real values of the variables and the real world observations).

A visual interpretation of the data-driven generation of a network model encoding the interaction between variables of the system by the Inatas System Modeller software [video]:



System Modeling
Data Analysis
Predictive Analytics
Automated Decision Making
Process Optimization

THE DATA PROCESSING

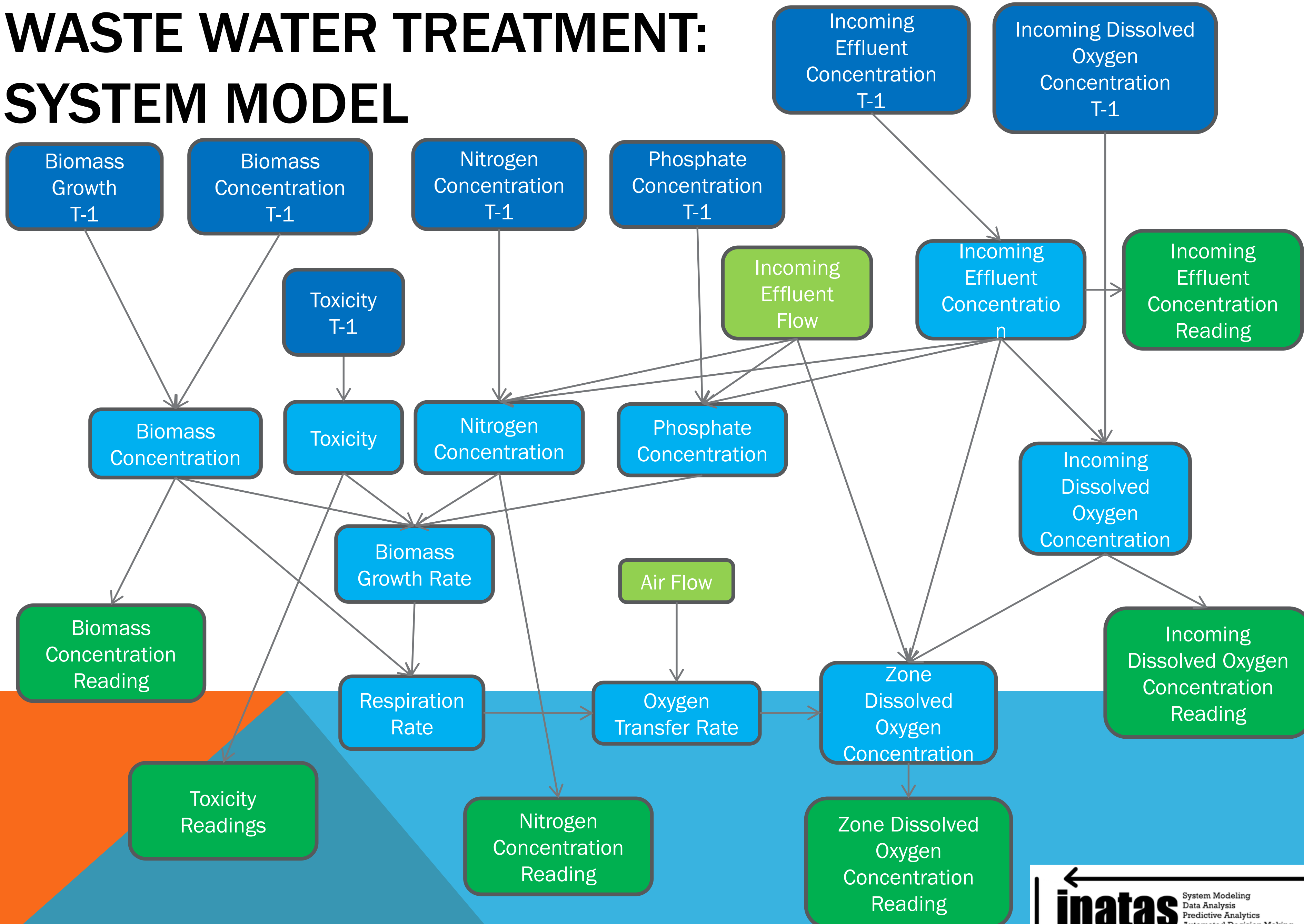
Automated learning algorithms produce high quality models of the relationships between the variables of the targeted system. This model can be built directly from existing data, avoiding difficult, expensive and time-consuming manual model design.

Expert knowledge about likely effects of unusual events is incorporated into the learning process.

Real time querying of the model provides simple, easy to understand output of the current knowledge of the level of phosphate.

Graphic lay-out of the model permits easy comprehension (and checking) of the relationships thought to be established in the system.

WASTE WATER TREATMENT: SYSTEM MODEL

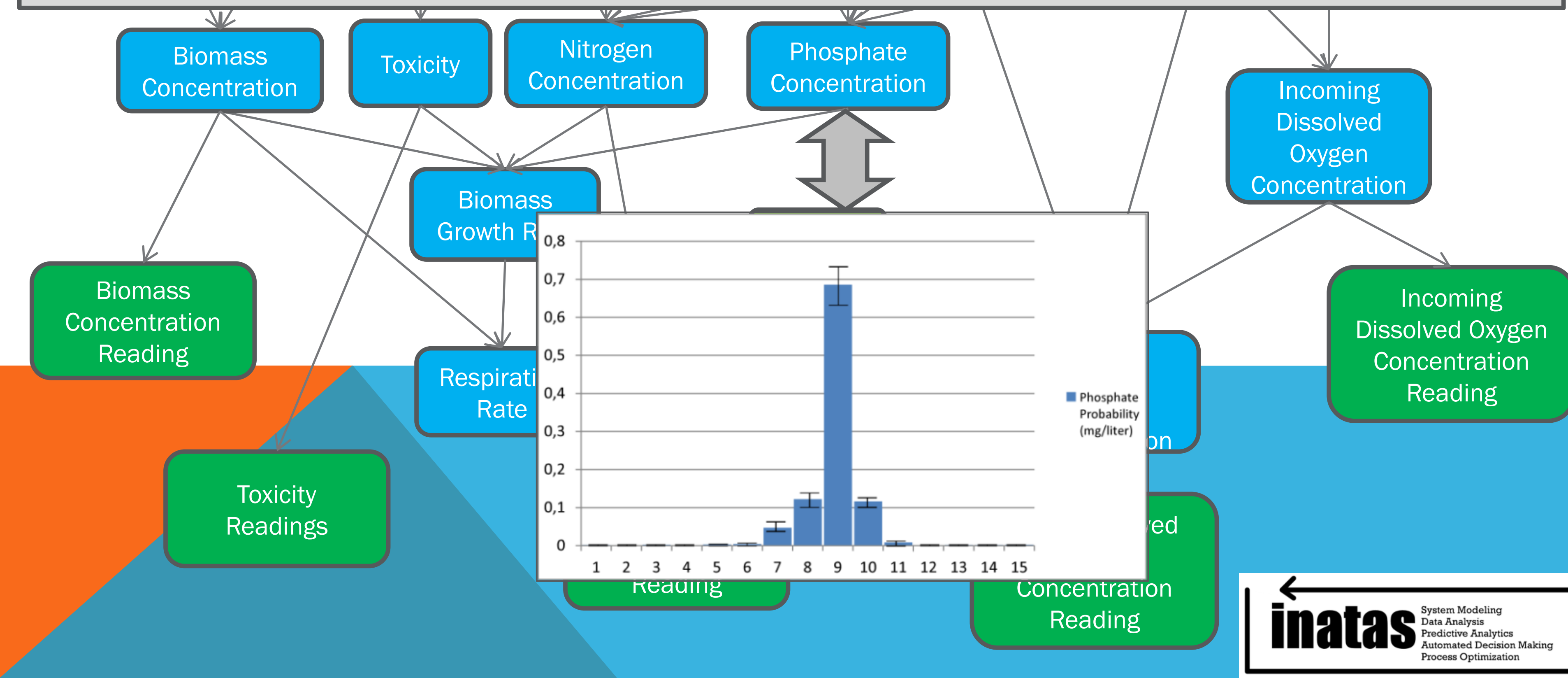


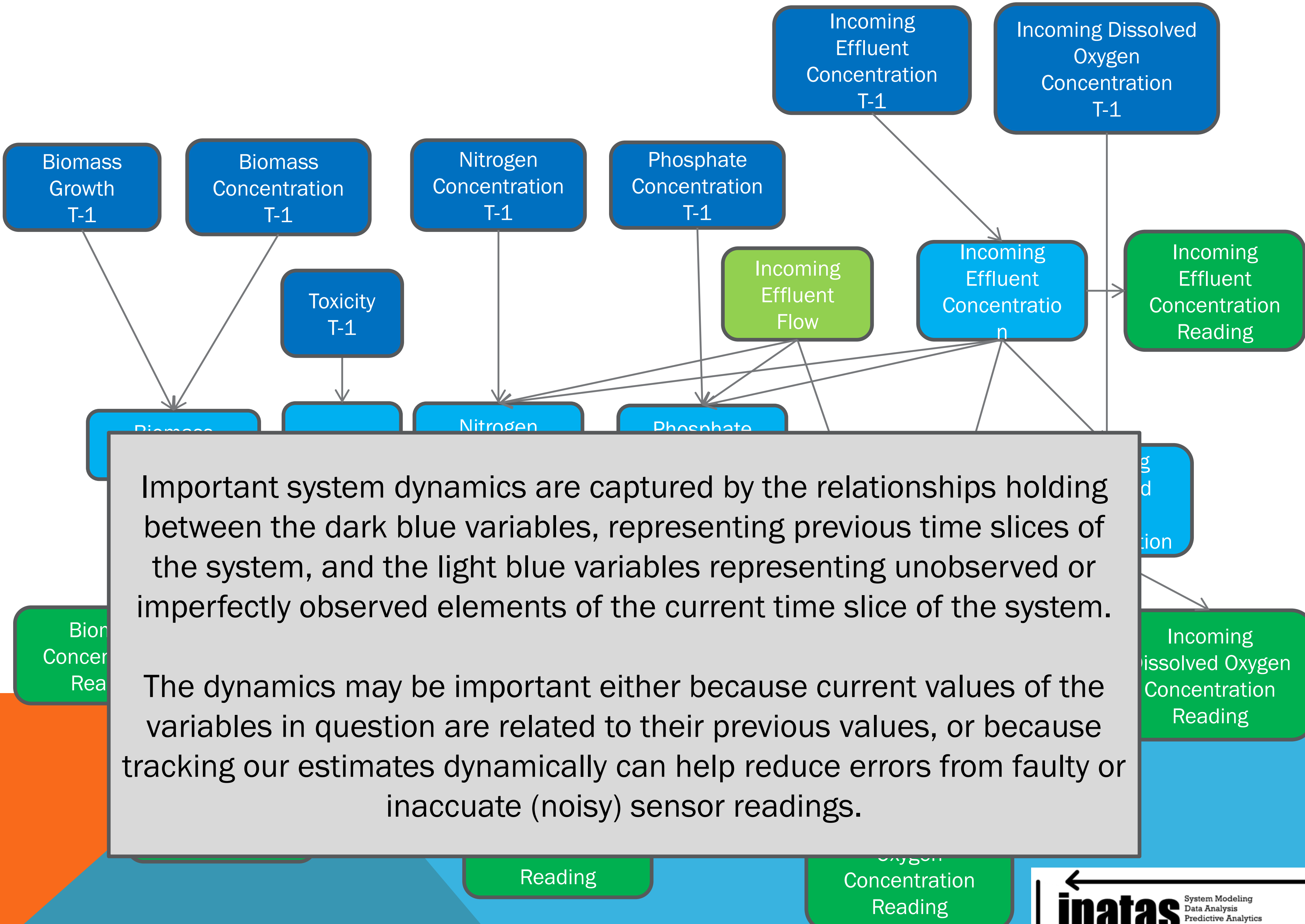
inatas
System Modeling
Data Analysis
Predictive Analytics
Automated Decision Making
Process Optimization

Relationships found to be present are encoded by the structure of the network. The nature of these relationships is given by conditional probability distributions – with each node is a conditional distribution for that variable, given its parents.

We perform calculations on this network of conditional distributions to provide estimates of the aposteriori distribution of unobserved variables – and in particular the current level of phosphate – given observed variables.

Given is a sample of what such an estimate of the aposteriori probability distribution for phosphate might look like. The bars on the columns are confidence intervals.



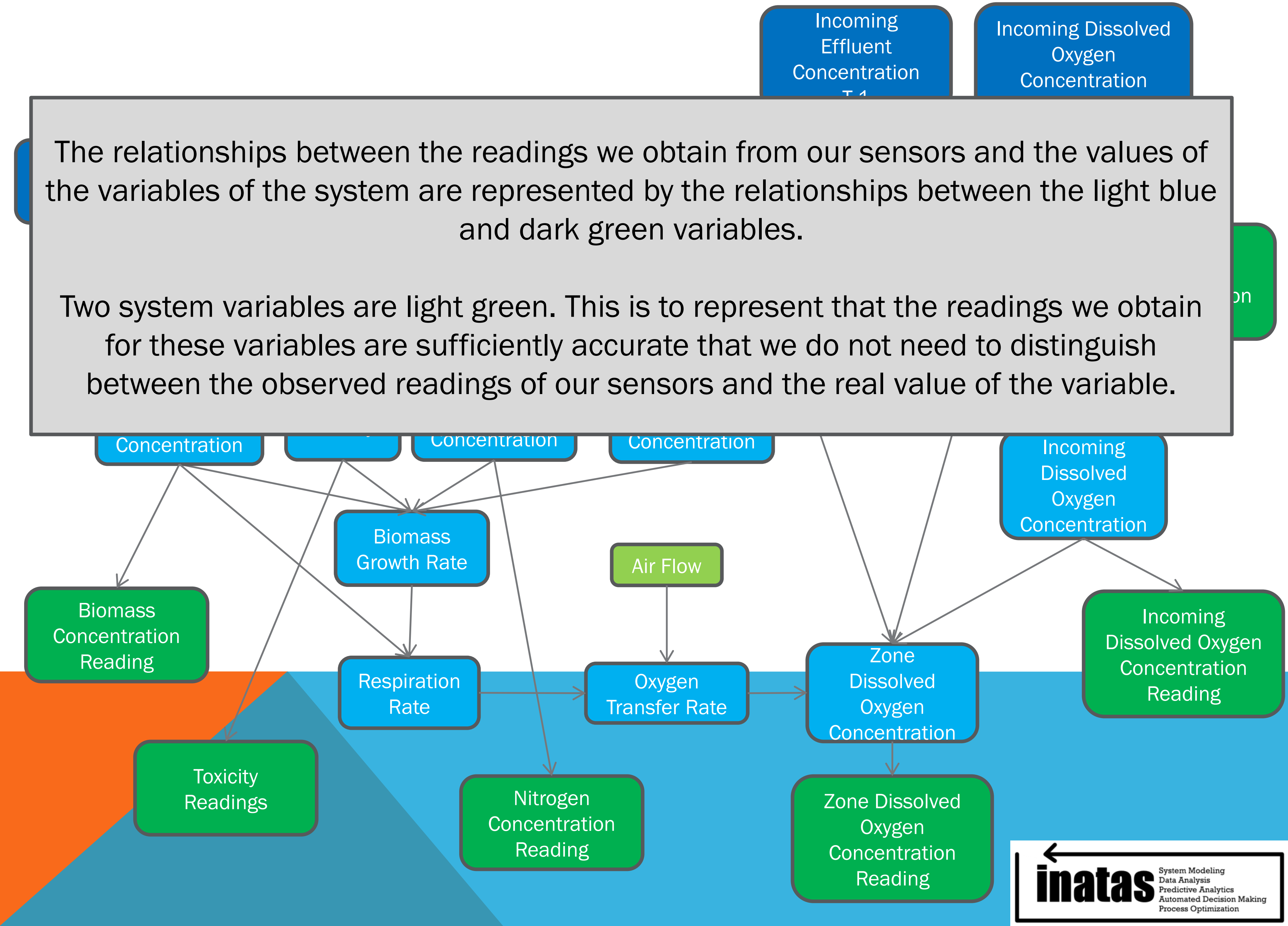


Important system dynamics are captured by the relationships holding between the dark blue variables, representing previous time slices of the system, and the light blue variables representing unobserved or imperfectly observed elements of the current time slice of the system.

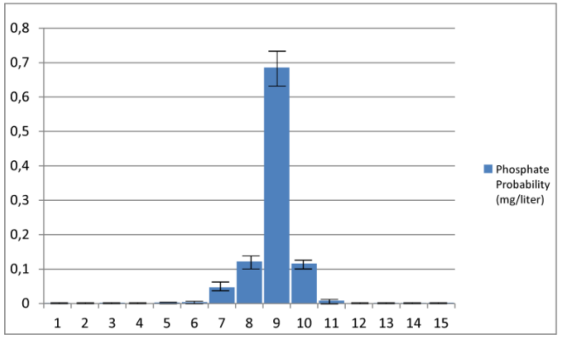
The dynamics may be important either because current values of the variables in question are related to their previous values, or because tracking our estimates dynamically can help reduce errors from faulty or inaccuate (noisy) sensor readings.

The relationships between the readings we obtain from our sensors and the values of the variables of the system are represented by the relationships between the light blue and dark green variables.

Two system variables are light green. This is to represent that the readings we obtain for these variables are sufficiently accurate that we do not need to distinguish between the observed readings of our sensors and the real value of the variable.



USES



Inexpensive phosphate tracking

- Highly accurate model
- Highly informative
  - Probability distributions, not point estimates
  - Confidence intervals
- Significantly cheaper than directly tracking the phosphate level.
- Completely automated, with inbuilt warnings for need for human interaction

Scenario testing option: We can simulate the effects on the phosphate level of the system by testing the model with particular values and examining the consquences.

It is possible to use the model for real time system control, by specifying the interval values by which the system's control should be set as to optimize the system output. See example in case 3.

inatas System Modeling Data Analysis Predictive Analytics Automated Decision Making Process Optimization

CASE 2

Capacity Planning/Demand Forecasting

DESCRIPTION OF THE PROBLEM/SITUATION

Effective capacity is the maximum amount of work that an organization is capable of completing in a given period due to constraints such as quality problems, delays, handling of material, product capacity , equipment, etc.

The company wishes to ensure that their effective capacity is sufficient for expected demand

The company has at their disposal large amounts of data, stored in databases. It is generally not fully exploited, and stakeholders have limited means of extracting actionable information from this data.

The company also has internal experts able to provide quality assessments of effects of isolated changes in their operating processes and market environment, based on years of experience.

HOW INATAS EXTRACTS NEW KNOWLEDGE

The software has automated procedures for discovering the variables which are directly relevant to the company's effective capacity and their expected demand.

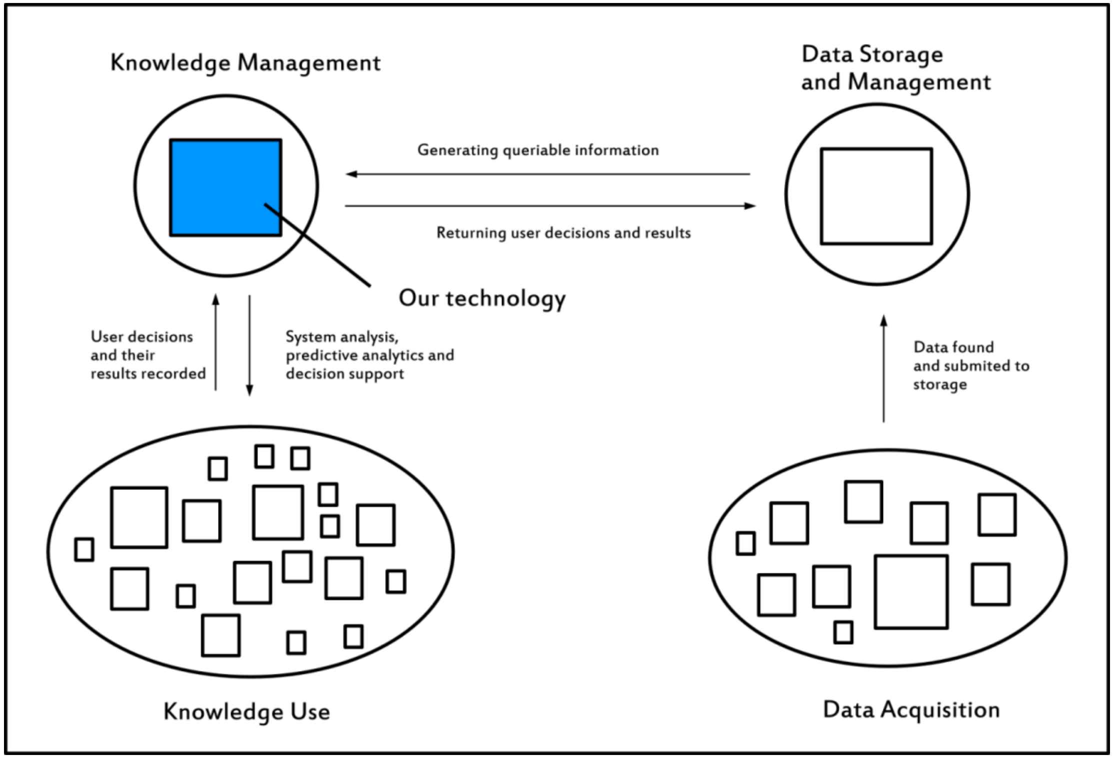
- Large numbers of variables are correlated with capacity and demand.
- Inatas' technology identifies those that are directly influential.

The software has partially automated processes to generating structured, queriable models from the unstructured or unqueriable data sources available and expert knowledge about specific facets of the system.

- Permits easy, holistic comprehension of the empirical and expert knowledge available.
- Resulting queriable and understandable models are usable in decision making and in stakeholder explanation.

The software allows for guidance or automation of decision-making based on the generated optimal policies.

HOW INATAS FIT INTO THE BIG PICTURE



inatas
System Modeling
Data Analysis
Predictive Analytics
Automated Decision Making
Process Optimization

RESULT: IMPROVED UTILISATION OF PRODUCTION RESOURCES

Data is accessible and queriable in the Inatas software.

Examine effects on effective capacity (at particular times) of changes in relevant variables. E.g.:

- Opening new/Closing old production facilities.
- Shortages or price spikes in required materials.
- Changes in sales price of output, where this makes certain production techniques cost (in)efficient.
- Transportation time.

Examine the effect of expected demand of changed market conditions. E.g.:

- Particular seasonal events.
- Long-term trends.
- Competitor/Regulator actions.
- Unexpected events (disasters, World Cup victories)

Quantify the probability of such changes occurring given current conditions

Note that the variables involved are dependent: In many cases, increased production capacity may affects the cost price of the outputs, affects the production policies of competitors, affects the prices of raw materials, etc. Inatas' analysis is able to deal with large systems of inter-dependent variables.

inatas System Modeling Data Analysis Predictive Analytics Automated Decision Making Process Optimization

USES IN POLICY GENERATION AND EVALUATION

The generated models:

- Generate and evaluate policies for maximizing the probability that effective capacity is capable of meeting demand, given the cost constraints.
  - These policies can be both general - taking into account the probability of all possible scenarios, and their possible effects on capacity – and conditional on particular events occurring.
  - These policies represent the optimal decisions given the empirical observations, specified expert knowledge and specified utility of outcomes.
- Permit full understanding of the variables that are under the company's control.
- Examine potential scenarios and their expected impact on the company's effective capacity and expected demand.
- Identify the areas which the company should focus upon to maximize their utilisation of the company's assets. This is particularly useful in deciding upon company focus and the development of KPI (key performance indicators).

inatas System Modeling Data Analysis Predictive Analytics Automated Decision Making Process Optimization

CASE 3

Manufacturing company

- Use of Inatas software for Quality Control
- Use of Inatas software for System Control

THE PROBLEM

The company produces a key component for the car industry. They have already a system in place for collected real-time data from the manufacturing processes.

They wish to utilize Inatas' software to:

- Minimize the amount of defective products produced as far as is cost effective to them (and their customers).
- Analyse the various condition of their processes (whether it is 'In-Control' or 'Out-of-Control').

THE USE OF DATA TO GENERATE NEW KNOWLEDGE

The company has historic data sets, combined with specifications made by company system experts about likely effects of unusual events, which we combine to produce the Inatas model.

A continuous, real time control loop is set up by the company:

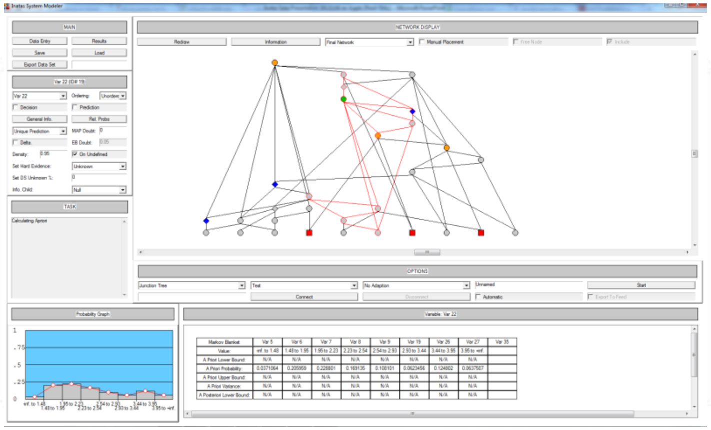
- Data, collected in real time, is transmitted from the manufacturing process to the Inatas software.
- The current state of the system is analysed by the model. Decision theoretic algorithms calculate settings for system elements that (i) minimize the expected number of defects given the current environment; and (ii) maximize the probability of the system staying 'In-Control'.
  - The company is able to specify the relative importance of these two, potentially conflicting, goals.
  - The optimal settings of variables that are under the system's control will depend on the values taken by 'environmental' factors. By 'environment' we mean the variables not under the system's control. For example, it may be that the temperature schedule of an annealing process that minimizes defects may depend upon the level of humidity. This distinction is not unchanging: The relationships discovered by the modelling process may lead to the company deciding to bring particular environmental variables under their control.

If the system is 'In-Control' and predicted to remain so, these settings are then implemented. If not, the system alerts the operators.

inatas System Modeling Data Analysis Predictive Analytics Automated Decision Making Process Optimization

HOW THE REAL TIME CONTROL LOOP WORKS

Current sensor readings passed from system sensors to model.



If the system is 'In-Control' and expected to stay so, the calculated optimal settings for defect minimization and control maximization are automatically implemented.

If the system is 'Out-of-Control' or expected to become so, human managers are alerted.

inatas
System Modeling
Data Analysis
Predictive Analytics
Automated Decision Making
Process Optimization

ADVANTAGES OF INATAS

Minimize number of defects.

Minimize the down-time of the manufacturing system.

Use of the proprietary Inatas data API to use and analyse real time data

CASE 4

Maintenance and fault detection in wind turbines

THE PROBLEM

The company produces wind turbines. They have a test turbine set up from which to collect data, but cannot equip turbines in use with full scope real-time sensor systems. Instead, data is collected on only a subset of variables on the deployed turbines. The sensors deployed are noisy (their measurements are not precise, and may face environmental interference).

They wish to utilize Inatas' technology to:

- Optimize their maintenance schedule: Servicing turbines only when required while minimizing the number of deployed turbines that experience faults.
- Obtain root cause analysis of faults that do occur.

DATA AVAILABLE AND METHOD

Data from the test turbine, combined with specifications by internal system experts with knowledge about likely effects of unusual events, are combined to produce the Inatas model.

This models describes:

- The influences between the elements/variables of the system.
- The dynamics of the system (how the current state is related to the previous state).
- A sensor model (how the readings of the noisy sensors are related to the real states of variables).

For deployed turbines, a monitoring system is deployed, which involved a real time loop:

- Data about the state of those variables which are tracked at the deployed turbines is transmitted to the Inatas model in real time.
- The network maintains an ongoing set of estimates of the values the variables are likely to be taking, given the sequences of values observed. (i.e. a dynamic state model is constructed and updated based on the new observations.)
- On the basis of the current estimates of the state of the system, decision theoretic algorithms calculate whether maintenance should be undertaken. If so, operators are alerted and the components of the system requiring maintenance are specified based on the data.

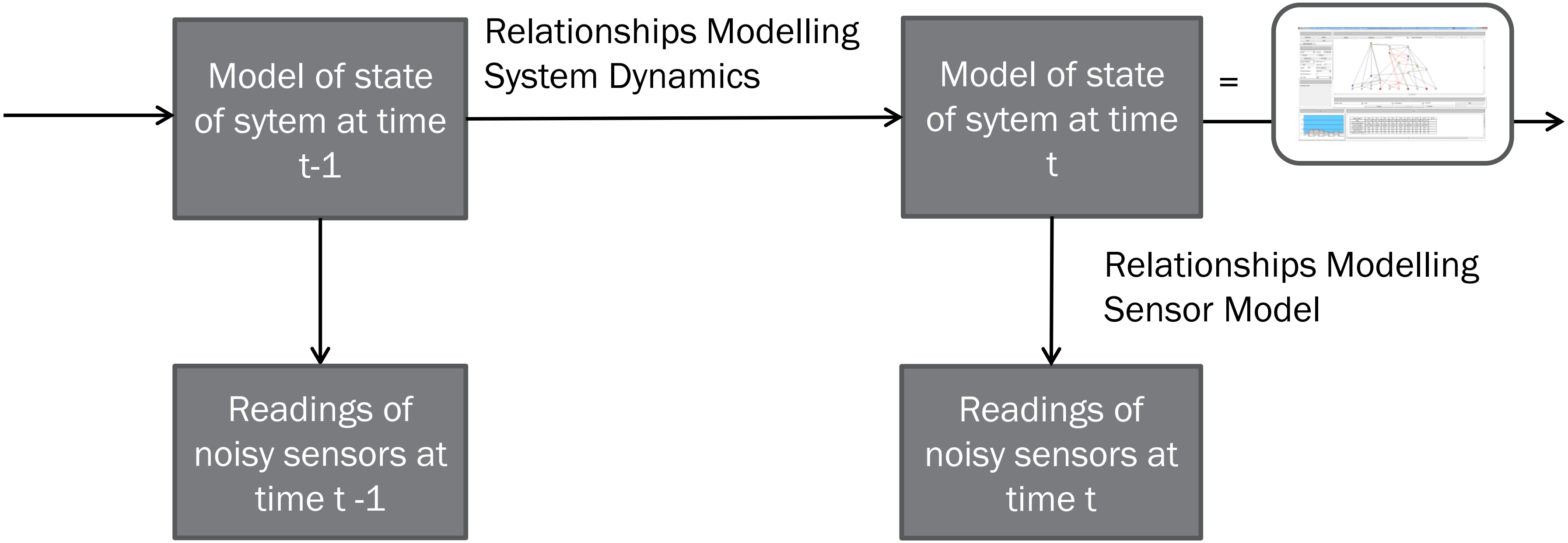
Should a fault be detected:

The model undertakes root cause analysis, calculating which components of the system are most likely to have caused the detected fault.

The wind turbine owners are alerted and a ranked set of hypotheses regarding the cause of the fault is specified.

inatas System Modeling Data Analysis Predictive Analytics Automated Decision Making Process Optimization

DYNAMIC MODELLING



When a fault is detected, probabilities are traced back through the network to provide valuable information regarding the likely causes, and, if useful, the likely time of these causes.

inatas
System Modeling
Data Analysis
Predictive Analytics
Automated Decision Making
Process Optimization

USES

Minimize unnecessary maintenance.

Minimize faults through timely maintenance.

Provide valuable information regarding the likely causes of detected faults.

CASE 5

Development of medical diagnostic tool

THE PROBLEM TO BE SOLVED

The company has planned proof-of-concept (PoC) clinical studies for its diagnostic platform

- The company's product is designed to be an objective diagnostic tool for measuring biomarkers for a disease, the disease intensity and therapeutic response for drug treatment
- The clinical studies involve quantitative measurements of a high number of biomarkers on patients as well as healthy persons, at several different time-points, on healthy and disease affected patients
- Biomarker measurements will be correlated to clinical – (e.g. disease score, treatment, etc.) and non-clinical (e.g. age, sex, habit) parameters of patients

The complexity of variables in the company's clinical studies require a multi-variant analyses to maximize the extraction of valuable data from the studies

inatas System Modeling Data Analysis Predictive Analytics Automated Decision Making Process Optimization

THE DATA SETS - AN ILLUSTRATION

Data from biomarker measurements from the skin of healthy persons and disease affected patients

- From 10 different biomarkers
- From 20 treated and 20 previously untreated patients
- From 4 different medical treatments
- From 5 different time-points
- From healthy person and disease affected patient
- From different sampling techniques/sites to obtain the sample

Data from clinical parameters

- Overall clinical score system
- Individually assessed scores at the individual sites of measurements: 5 parameters measured
- An additional 10 parameters such as concomitant diseases, time of diagnosis, length of disease etc.

Data from non-clinical parameters

8 parameters such as age, sex, smoking habit, weight etc.

inatas System Modeling Data Analysis Predictive Analytics Automated Decision Making Process Optimization

WANTED OUTCOME

From the company's first clinical studies in patients and healthy volunteers the company would like to learn:

- Are there quantitative correlations between individual biomarkers and clinical parameters, during diagnosis and therapeutic response?
- Are measurements of individual, or groups of, biomarkers affected by individual, or groups of clinical and non-clinical parameters?
- What - if any- transformations of data (linear, logarithmic etc.) could improve the correlation among the above data

For this we would like to extract:

- Correlations between individual biomarkers and individual clinical parameters
- Correlations between groups of biomarkers and groups of clinical parameters

USING INATAS' TECHNOLOGY FOR ANALYSIS

Step 1:

- A number of mutual information analyses are performed to examine whether transformations of the involved variables is likely to result in improved results.

Step 2:

- Any background knowledge about the relationships that hold between the variables involved is entered in the form of structural constraints on the network, or on Bayesian priors on the parameters of the network.

Step 3:

- Data from the clinical trials is used to generate a model of the variables that specifies the relationships of influence that the data gives us reason to believe exist.
  - Many-to-many, many-to-one, one-to-many and one-to-one relationships are discovered in the modelling process.
  - While all correlations are reported, the principle relationships encoded in the network are much more informative than mere correlation. The model encodes the conditional independencies found. Two variables are often correlated while being conditionally independent because both are related to a third variable in one of numerous ways.
  - Knowledge of the conditional independencies present in a system provides information about the causal relationships present (see step 5).
  - Second order probabilities are used to inform the company about the confidence they should have in each discovered relationship.
  - In exceptional cases where multiple hypotheses for how the variables influence one another are comparably plausible, analysis over all such hypotheses indicates which relationships they share, which are present under some hypotheses and not other (and the relative probability of the models in which they occur and are absent), etc.

Step 4:

- The strength of various many-to-one and many-to-many relationships is measured by systematic querying of the model.

Step 5:

The information regarding conditional independencies found between the variables is combined with causal manipulation tests for causal analysis.

CASE 6

Bus scheduling: Allocation of buses

PROBLEM

This case will illustrate how Inatas software can be used for:

1. Optimal allocation of available buses to routes, for each time/day/season.
   - Day/time/season defined as desired by company.
2. Optimal re-allocation of buses to routes for special events, for each time/day/season.

There are other uses for Inatas technology in this area, including in maintenance scheduling and stress/scenario testing.

These will not be covered here.

THE OPTIMAL ALLOCATION OF BUSES - PROCESS

A 'layout' is defined as a specific time, day, season and route.

Step 1: Model relationships between available buses and customer satisfaction.

For each layout we need to model the relationships between the number of buses made available, observed passenger numbers, intermediary variables that are known to effect customer satisfaction and customer satisfaction.

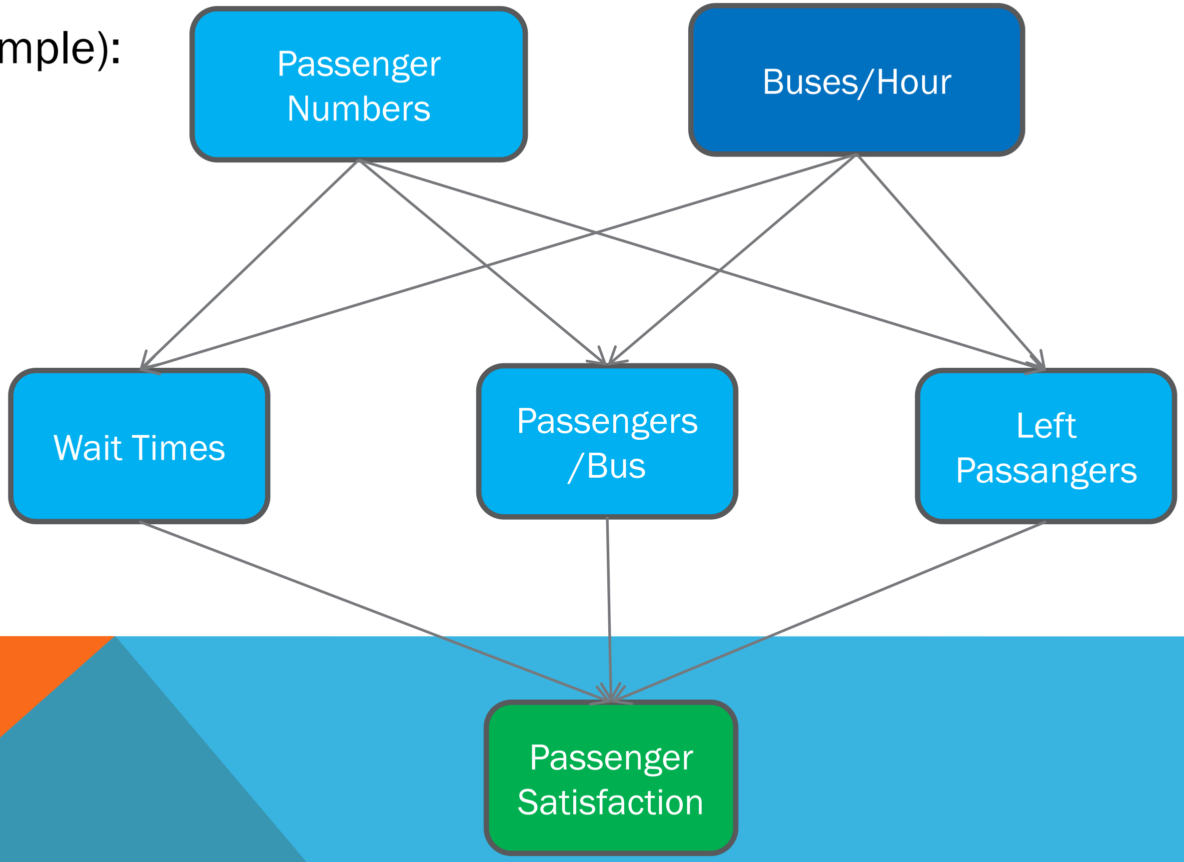
On the following pages the variables we wish to decide upon are shown in dark blue. The variables we wish this decision to maximize are shown in green, and other variables in light blue.

Note that we could have multiple variables to decide upon and maximize.

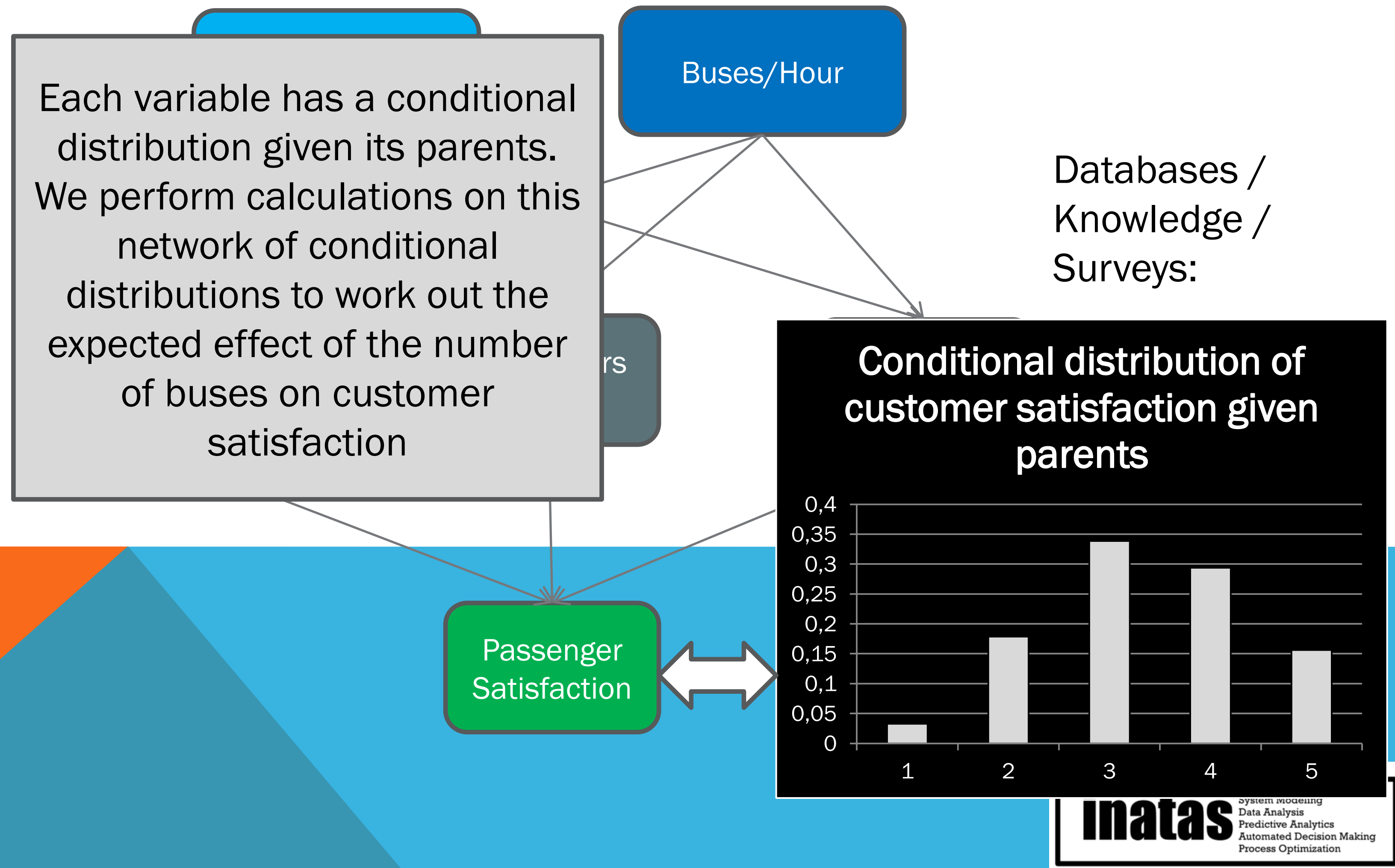
Specified events are only illustrative examples, based on our current understanding.

EXAMPLE MODEL OF THE RELATIONSHIP BETWEEN AVAILABLE BUSES AND CUSTOMER SATISFACTION FOR A GIVEN LAYOUT.

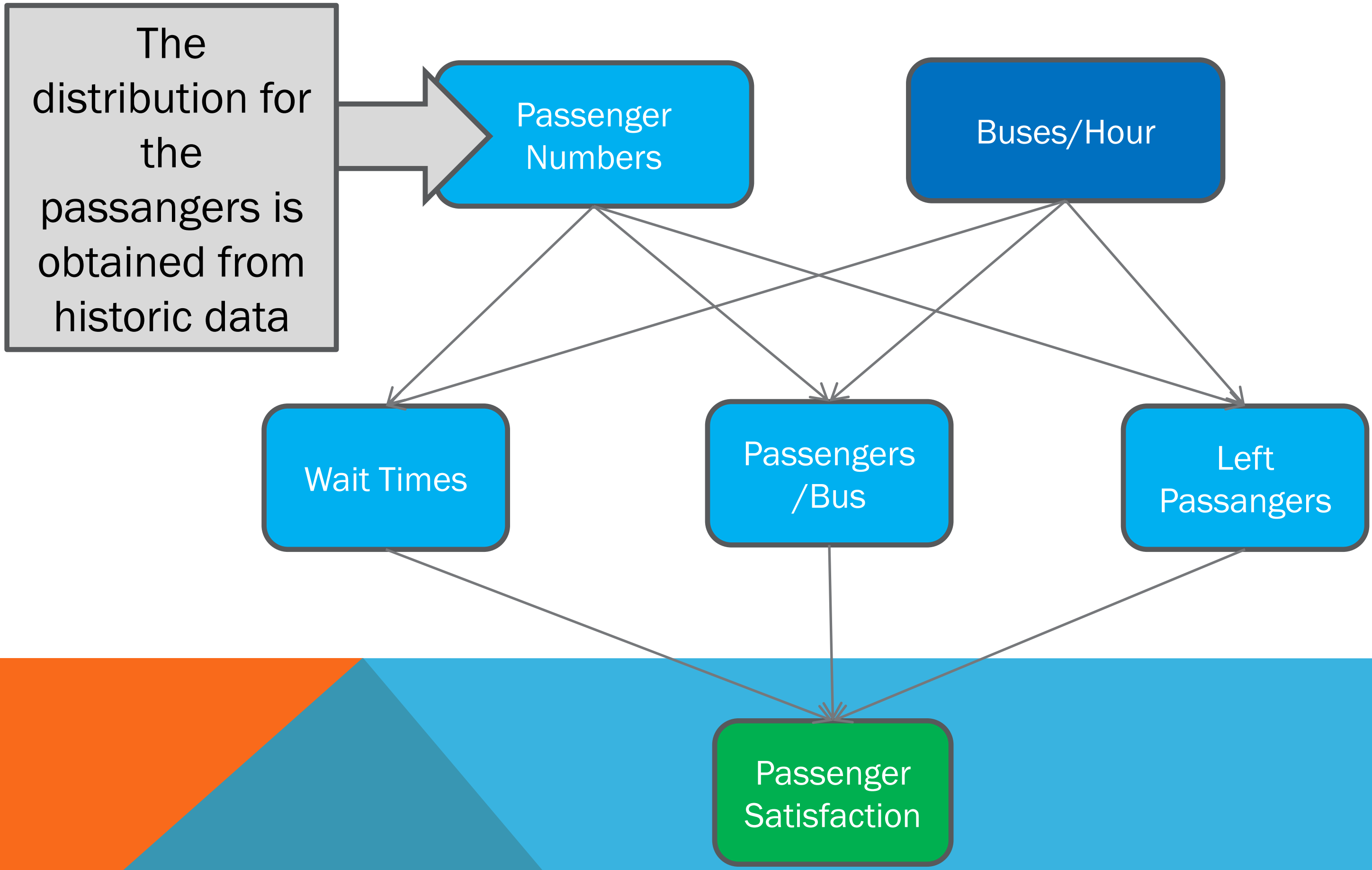
Model (example):



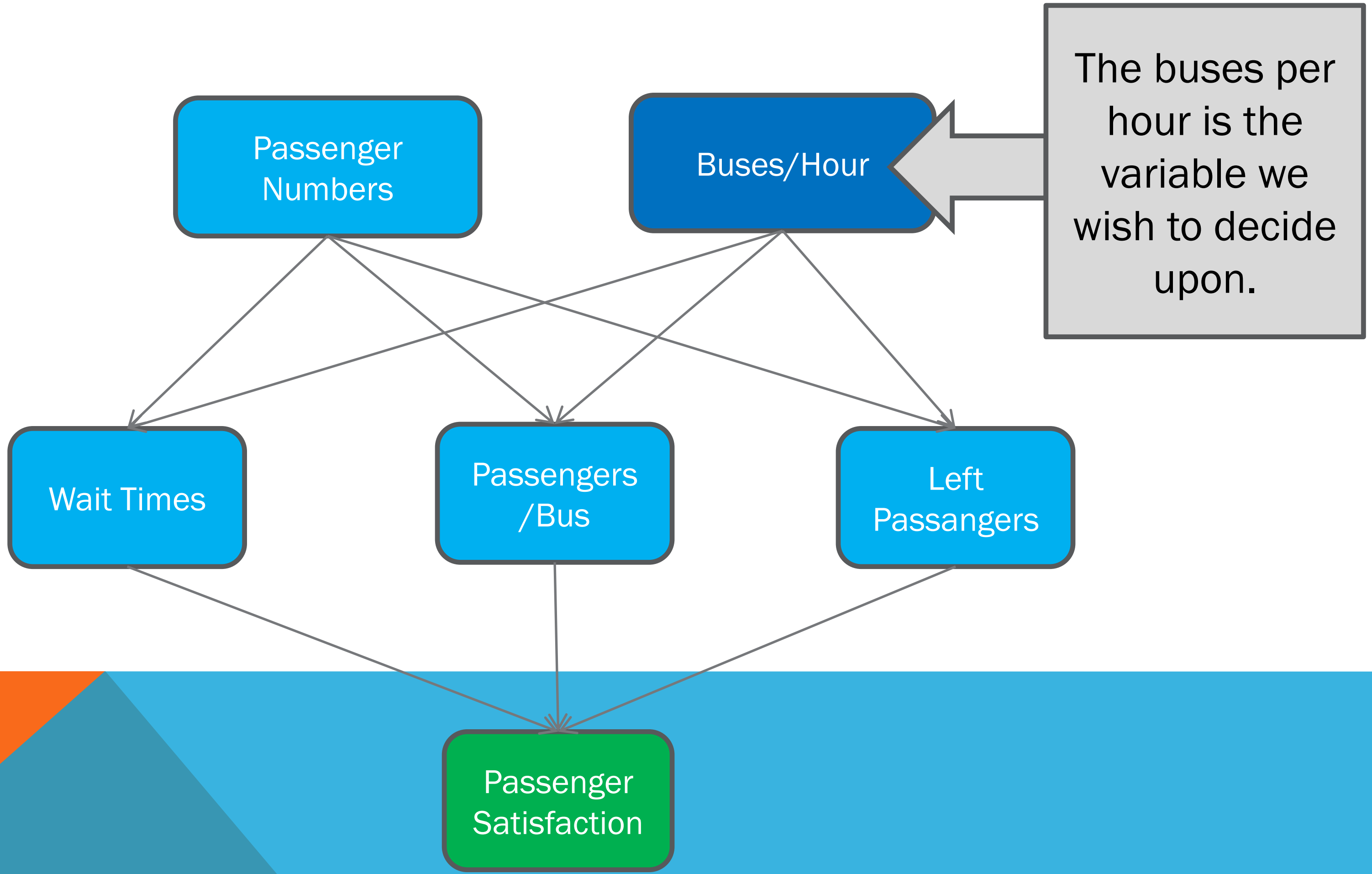
EXAMPLE MODEL OF THE RELATIONSHIP BETWEEN AVAILABLE BUSES AND CUSTOMER SATISFACTION FOR A GIVEN LAYOUT.



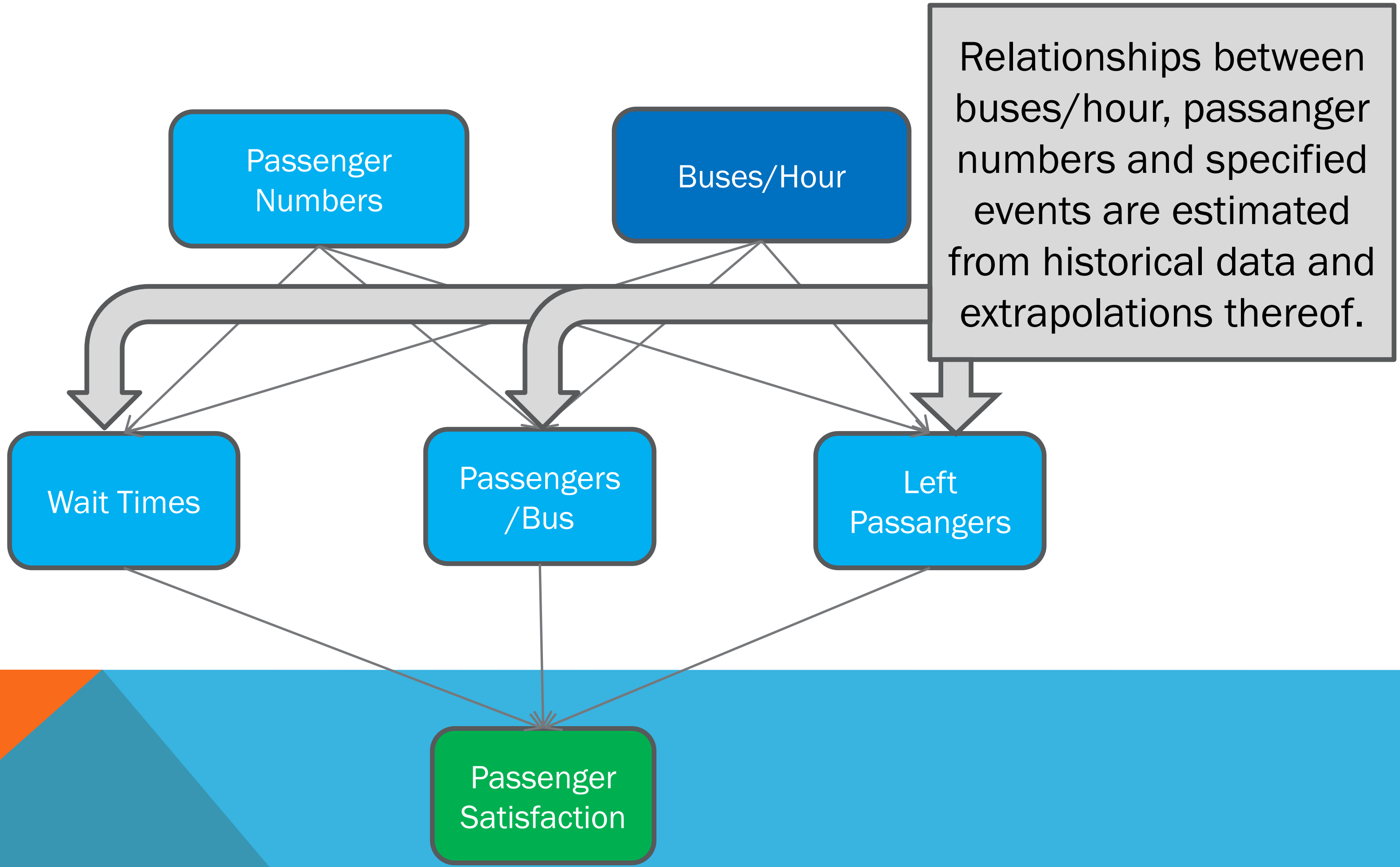
EXAMPLE MODEL OF THE RELATIONSHIP BETWEEN AVAILABLE BUSES AND CUSTOMER SATISFACTION FOR A GIVEN LAYOUT.



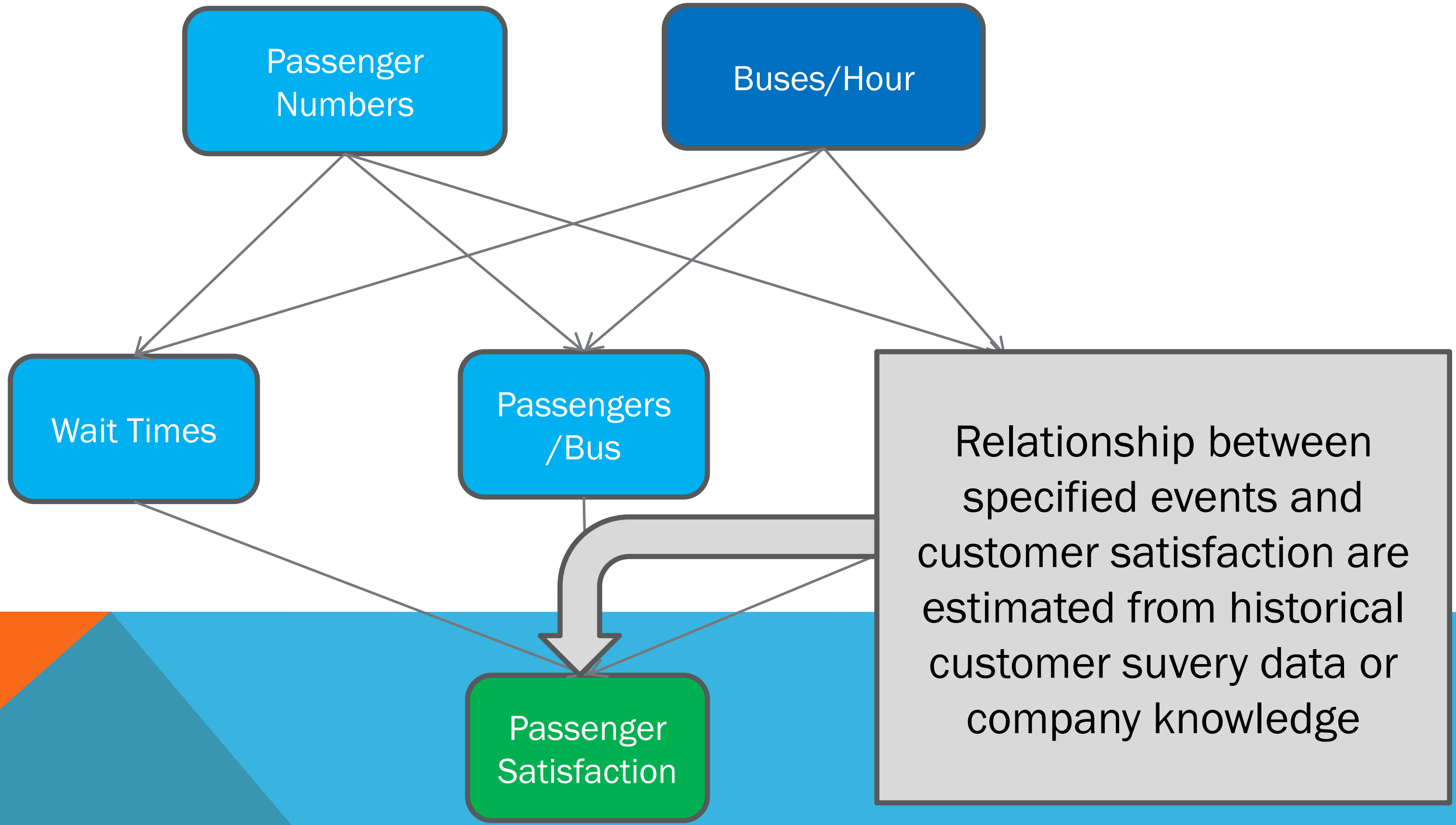
EXAMPLE MODEL OF THE RELATIONSHIP BETWEEN AVAILABLE BUSES AND CUSTOMER SATISFACTION FOR A GIVEN LAYOUT.



EXAMPLE MODEL OF THE RELATIONSHIP BETWEEN AVAILABLE BUSES AND CUSTOMER SATISFACTION FOR A GIVEN LAYOUT.



EXAMPLE MODEL OF THE RELATIONSHIP BETWEEN AVAILABLE BUSES AND CUSTOMER SATISFACTION FOR A GIVEN LAYOUT.



THE OPTIMAL ALLOCATION OF BUSES

Step 2: Provide a total satisfaction estimate for the layout.

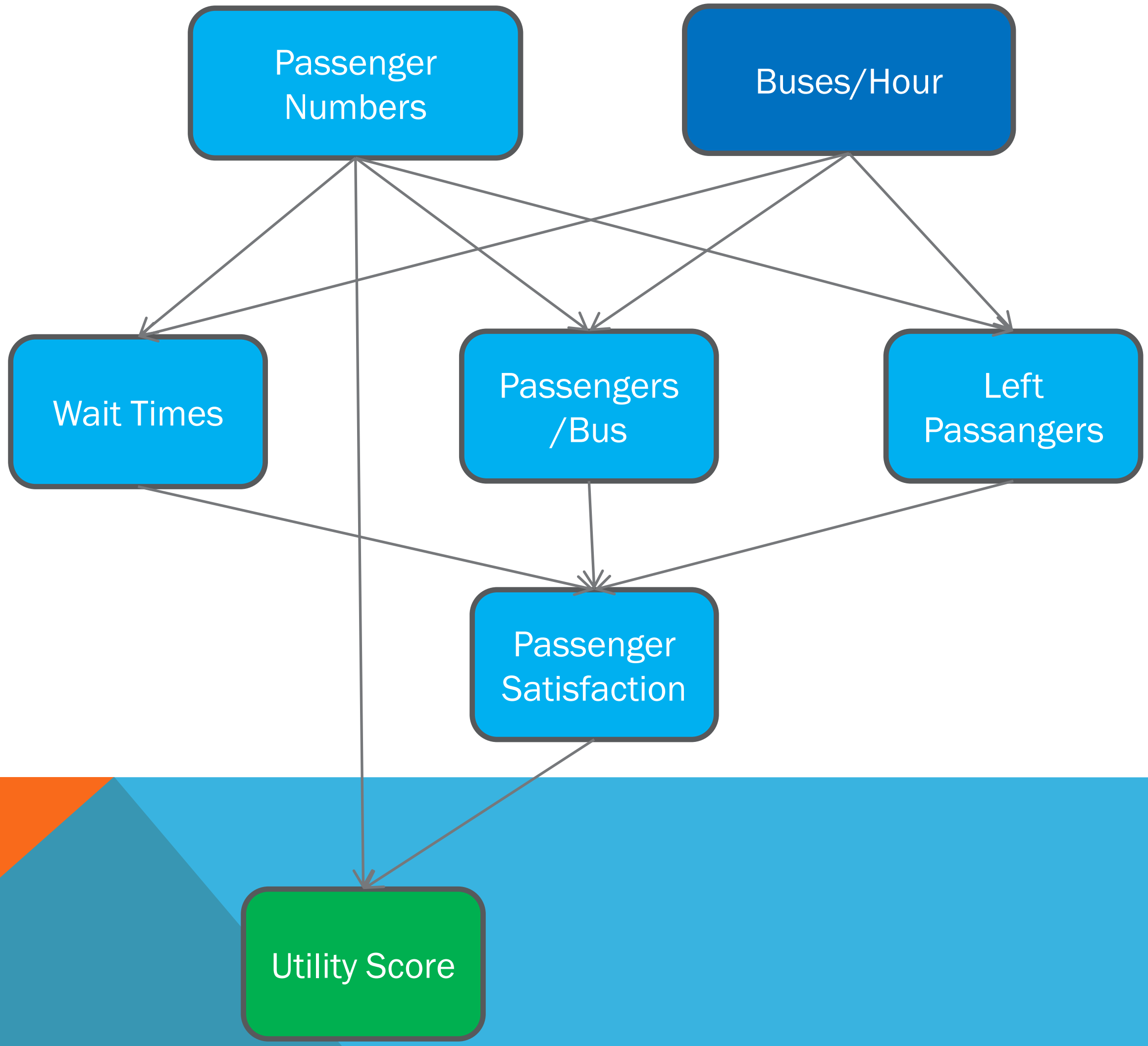
So far we know the relationship between the number of buses available to a layout and the distribution of customer satisfaction.

We also need to take into account the number of customers involved.

We now introduce a utilty score that is a function of the number of passengers and the customer satisfaction variables. If customer satisfaction is scored 1-5, a simple example would be the expected total of the shifted customers customer satisfaction scores by -2.

- Satisfaction is shifted from -2 to 2, and we calculate the expected total.
- Note that this is not a trivial result of the expected shifted satisfaction multipled by the the expected number of customers, since these two variables are dependent.

MODEL WITH UTILITY VARIABLE INCLUDED



THE OPTIMAL ALLOCATION OF BUSES

Step 3: Provide an opportunity cost.

We expect satisfaction to improve as the number of buses increases.

On the other hand we have only a finite number of buses that can be allocated at any given time.

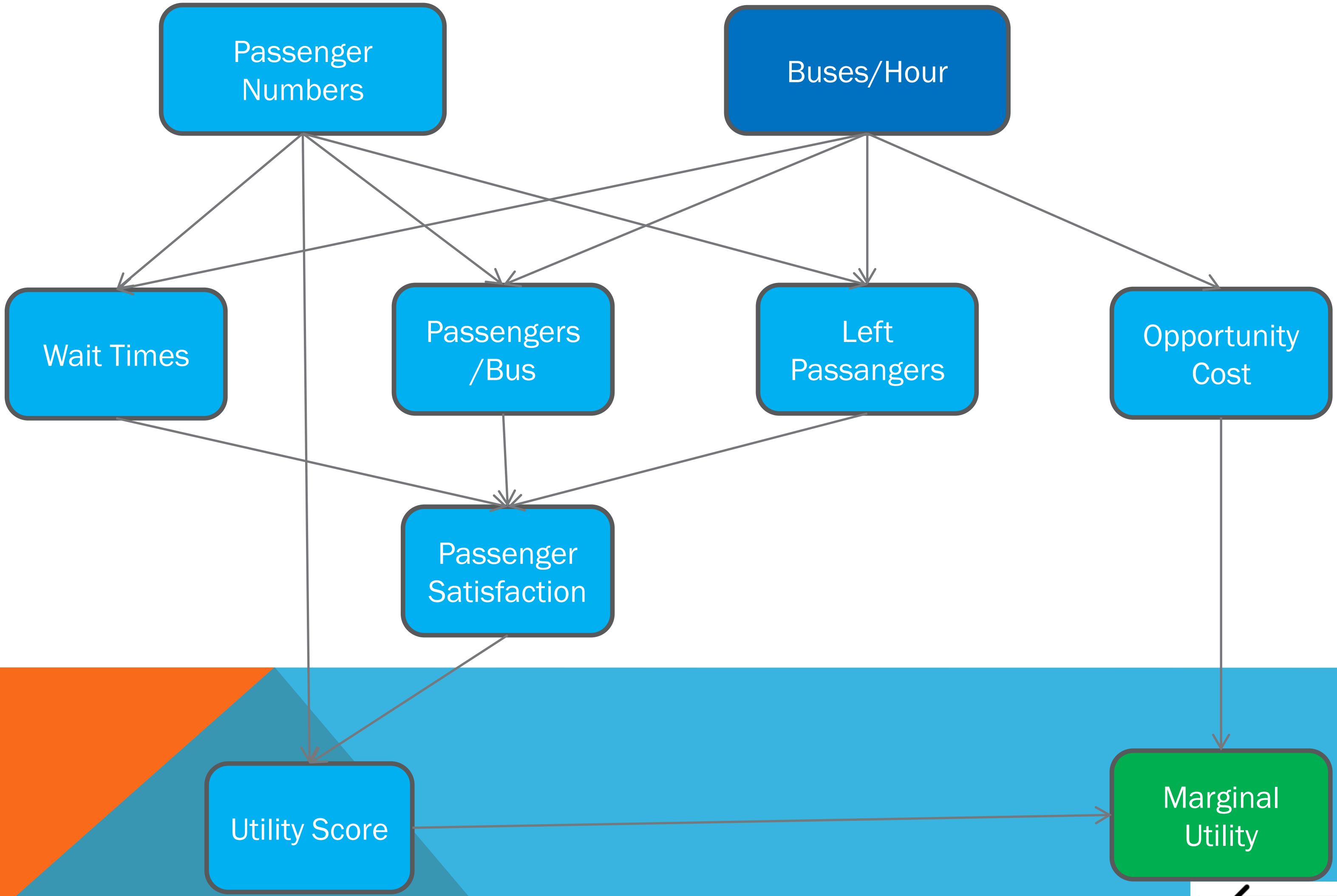
We need to take into account the opportunity cost of allocating a bus to a particular layout.

The opportunity cost of allocating *n* buses to a layout is the maximum utility that can be produced by allocating these *n* buses to other layouts that share the same time, day and season values.

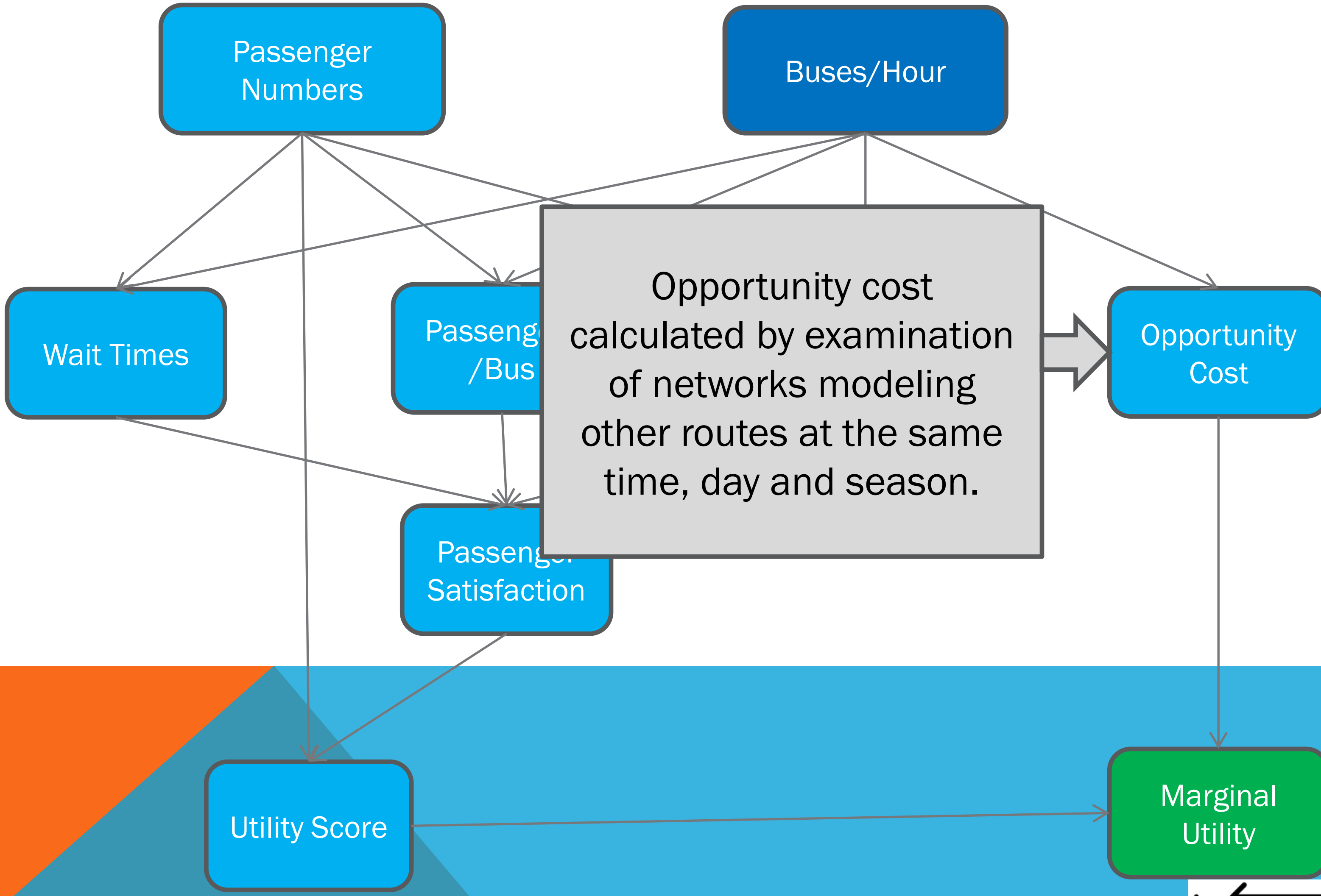
This can be calculated from the models generated by Inatas' software for these other layouts.

To calculate the optimal allocation of buses, we use the decision theoretic algorithms included in Inatas' software to maximize the marginal utility of all layouts, where this is the utility minus the opportunity cost.

MODEL WITH OPPORTUNITY COST AND MARGINAL UTILITY VARIABLES INCLUDED



MODEL WITH OPPORTUNITY COST AND MARGINAL UTILITY VARIABLES INCLUDED



Opportunity cost calculated by examination of networks modeling other routes at the same time, day and season.

OPTIMIZATION OF THE BUSINESS IS POSSIBLE

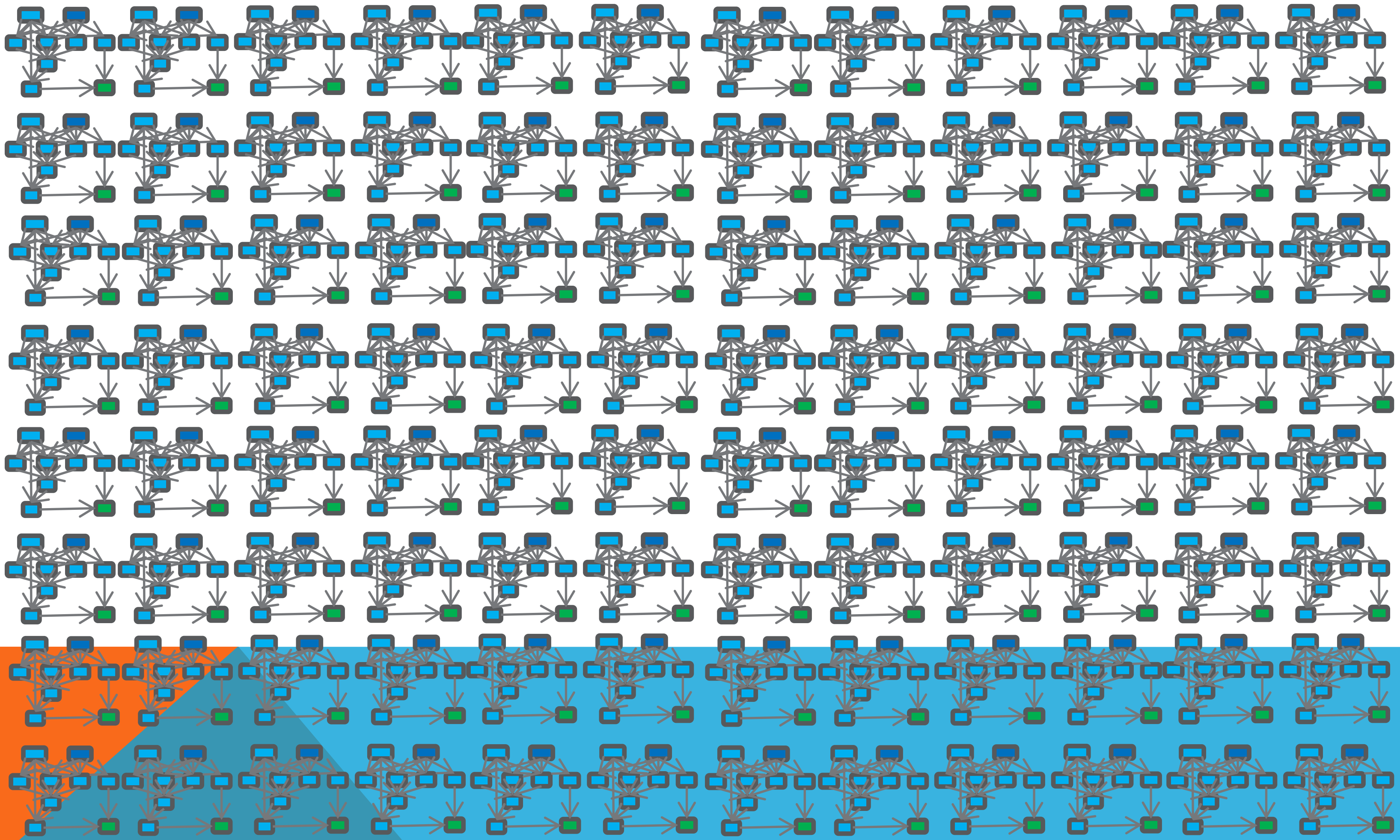
Using Inatas' software you are able to obtain the optimal allocation of buses for each route at each time, day and season.

But there are other "real life" issues to be solved:

- One example: What should we do when special events occur?

These can also be addressed ….

ZOOMING OUT TO VIEW MULTIPLE LAYOUTS…



inatas
System Modeling
Data Analysis
Predictive Analytics
Automated Decision Making
Process Optimization

REALLOCATING BUSES FOR AN EVENT

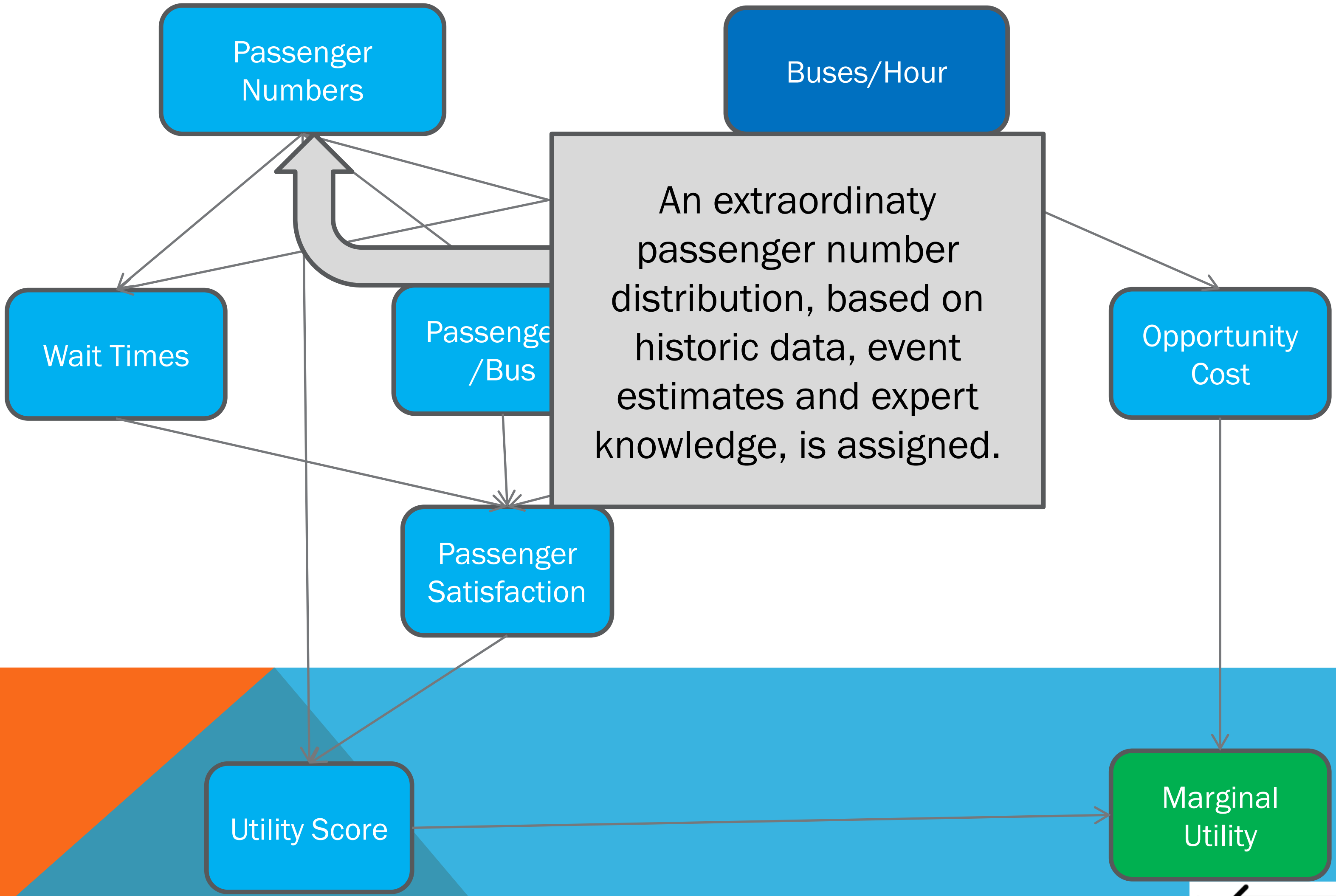
Certain layouts will periodically require additional buses due to extraordinary events.

We can use the models developed to calculate the optimal reallocation of buses to these layouts.

All that is required is that the layouts affected by the event in question be given event-specific passenger number distributions.

Once that is done, the interconnected set of networks can be used obtain the optimal allocation of buses for each route given the event occurring as described.

MODEL WITH OPPORTUNITY COST AND MARGINAL UTILITY VARIABLES INCLUDED



inatas
System Modeling
Data Analysis
Predictive Analytics
Automated Decision Making
Process Optimization

DEFINING YOUR CASE

Inatas' software is adaptable to providing solutions to many types of optimization

As an example the product link up to your databases and can extract data from these

Also the product can be used to process real-time data, as an example for decision support

An integral part of Inatas is our services to train customers in the use of our software and to impart an understanding of the models being used.

We work hard to develop the relevant model together with the clients, maximizing the advantages their domain knowledge can bring.

inatas
System Modeling
Data Analysis
Predictive Analytics
Automated Decision Making
Process Optimization

WHY INATAS' APPROACH IS SUPERIOR

Scalable: The system is able to scale up – you can predict multiple variables in very large systems.

Cost effective: Major savings of time and money obtained through the automated generation of advanced system models.

Fast: Processing of date is fast, based on the coding platform of the software. This is very relevant for real time system control of complex systems.

Informative: Provides a great deal of information about the relationships being analysed as well as information regarding the confidence you should have in the models doing the analysis.

On-line: Possible to use on-line data for analysis, and to get on-line advise based on live data coming in from the system.

inatas
System Modeling
Data Analysis
Predictive Analytics
Automated Decision Making
Process Optimization

CONCLUSION

Inatas' data analysis technology and software allows you to get and learn more from the data and expertise you have, and to understand, predict and control the systems you work with.