Diagnosis on a Principle Environmental Control System

- a study on improved functionality for fault monitoring in Gripen

Master Thesis
Reg nr: ISY-LiTH-EX-3067

Department of Electrical Engineering
Linköping University

Johan Nilsson

Avdelning, Institution
Division, department

Department of Electrical Engineering

Datum
Date

2000-06-06

Språk
Language

☐ Svenska/Swedish
■ Engelska/English

☐ _____

Rapporttyp
Report: category

☐ Licentiatavhandling
■ Examensarbete
☐ C-uppsats
☐ D-uppsats
☐ Övrig rapport

☐ _____

ISBN

ISRN

Serietitel och serienummer
Title of series, numbering

ISSN

LiTH-ISY-EX- 3067

URL för elektronisk version

http://www.fs.isy.liu.se

Titel
Title

Diagnosis on a principle Environmental Control System
-a study on improved functionality for fault monitoring in Gripen

Författare
Author

Johan Nilsson

Sammanfattning
Abstract

This thesis is carried out at the business unit Gripen of Saab Aerospace within the section for Thermal Analysis and System Simulation of General Systems. The work deals with automatic fault detection with focus on the Environmental Control System (ECS). The ECS main tasks are pressurization of cabin and cooling of avionics.

Automatic fault detection is traditionally performed mainly by using limit checking of measured signals and setpoint errors. An approach with a number of drawbacks such as difficulty in handling system transients and no systematic way for isolation of faults. Increasing demands on safety, reliability and economy has caught a growing attention for new diagnosis approaches. Model based diagnosis is found to be an approach with potential to increase performance and adding desired functionality.

A principle model of the ECS is developed for design and test of a model based diagnosis system. Not only the ECS is modeled but also relevant faults, some faults well known from long experience of the system and some hypothetical faults for the exemplification of diagnosis methods. The benefits of a model based diagnosis approach is compared to diagnosis methods implemented in the ECS today.

Nyckelord
Keywords

model based diagnosis, supervision, fault monitoring, FM, fault detection and isolation, FDI, Environmental Control System, ECS, JAS39 Gripen

Diagnosis on a Principle Environmental Control System

- a study on improved functionality for fault monitoring in Gripen

Master Thesis

Department of Electrical Engineering
Linköping University

Section for Thermal Analysis and
System Simulation of General Systems,
Gripen, Saab

Johan Nilsson

Supervisor: **Birgitta Lantto**
Erik Frisk
Examiner: **Lars Nielsen**
Linköping, 21st May 2000

Abstract

This thesis is carried out at the business unit Gripen of Saab Aerospace within the section for Thermal Analysis and System Simulation of General Systems. The work deals with automatic fault detection with focus on the Environmental Control System (ECS). The ECS main tasks are pressurization of cabin and cooling of avionics.
Automatic fault detection is traditionally performed mainly by using limit checking of measured signals and setpoint errors. An approach with a number of drawbacks such as difficulty in handling system transients and no systematic way for isolation of faults. Increasing demands on safety, reliability and economy has caught a growing attention for new diagnosis approaches. Model based diagnosis is found to be an approach with potential to increase performance and adding desired functionality.
A principle model of the ECS is developed for design and test of a model based diagnosis system. Not only the ECS is modeled but also relevant faults, some faults well known from long experience of the system and some hypothetical faults for the exemplification of diagnosis methods. The benefits of a model based diagnosis approach is compared to diagnosis methods implemented in the ECS today.

Acknowledgments

I would like to thank my supervisor Erik Frisk for excellent guidance and many interesting discussions. Birgitta Lantto is greatfully acknowledged for her support and positive spirit throughout the work. Thanks also to the crew at GDGT for giving me a good time at Saab with many laughters and tasteful coffee breaks.

Contents

Some notations used

In general

F - fault mode
G(s) - model transfer function
ΔG(s) - model error transfer function
H - hypothesis test
J - threshold
R - set of fault modes
S - diagnosis statement
T - test quantity
y - measured signal
Ω - set of all fault modes

Quantities in the principle model

a - valve position [rad]
A - orifice effective open area [m^2]
M - mass flow [kg/s]
P - absolute pressure [kPa]
T - temperature. [K]
V - volume [m^3]

Subscripts refering to different parts of the model

amb - ambient
ao - avionics outlet
av - avionics
cab - cabin
cin - inlet of cooling pack
co - cabin outlet
cp - cooling pack
eng - engine, or first part of the principle model
he - heat exchanger
15, 16, 18, 22 - valves, corresponding to Saab notation

Abbreviations

ECS - Environmental Control System
FM - Fault Monitoring
SC - Safety Check
RLS - Recursive Least Squares
FDI - Fault Detection and Isolation
MFL - Manual Fault Localization (Isolation)

1 Introduction

Saab Aerospace is a business area within Saab AB. The main enterprise is the development and production of the Gripen fighter. Gripen is a fourth generation aircraft, which refers to an extended use of integrated computerized systems. Information are provided from all parts of the aircraft, which opens new possibilities to analyze aircraft conditions. In this thesis, model based diagnosis is used to evaluate available information with purpose to extract system condition. The work has been performed at the section for system simulation and thermal analysis of general systems.

1.1 Objectives

The objective with this thesis is to work with model based diagnosis on a general aircraft system in Gripen. The main task is to exemplify diagnosis concepts and design process by building a model based diagnosis system. Another important aim is to take an inventory of diagnosis methods implemented today. An overview of functionality and methods implemented today is obtained by reading documentation and talking to people involved in the subject. Last of all, the work should sum up to a discussion and reflections on possibilities with model based diagnosis applied to a general aircraft system.

1.2 Background

The general aircraft systems are characterized by being large, dynamic and nonlinear. Traditionally these systems are supervised with sensor redundancy or limit and trend checking. Increased requirements on reliability, safety and economy have opened the interest for new diagnosis methods. Model based diagnosis is an interesting approach to investigate further. This thesis is a first step to explore the possibilities with model based diagnosis on the aircraft systems.

A focus is held on the Environmental Control System, which originate is developed by a subcontractor and integrated in the aircraft. Increasing demands on performance and economy have made Saab take over the development of the system. This opens the possibility to add new functionality such as model based diagnosis.

1.3 Outline

Chapter two is a summary of the physics for the components used when building the model. In chapter three a principle model of the Environmental Control System is built. Chapter four deals with the building of a diagnosis system on the principle model and presents some results from simulations with the diagnosis system. In chapter five a short discussion with conclusions and ideas for further work is presented.

2 Model based Diagnosis

Traditionally diagnosis has been performed mainly by limit checking. When for example a sensor signal leaves its normal operating range, an alarm is generated. The normal range is predefined by using thresholds. This approach has some limitations especially in the case with highly nonlinear systems when the thresholds must be chosen according to a worst case scenario or tabulated for different operating conditions. Another disadvantage is the difficulty in isolating a present fault. There is no natural way of handling such a problem formulation since knowledge of how different faults affect the process is not built into the diagnosis system.

Another traditional approach is the use of hardware redundancy, where components are duplicated or even triplicated. There are at least three major drawbacks with this approach, hardware is expensive, it requires space and adds weight to the system. Advantages are a reliable diagnosis system with fast response.

As an alternative to traditional approaches, model based diagnosis have shown to be useful both as a complement and sometimes on its own. Compared to traditional approaches, model based diagnosis has a large potential to have the following advantages:

- It can be performed over a large operating range.
- Isolation of different faults becomes possible.
- It can provide higher diagnosis performance, for example smaller faults can be detected and the detection time is shorter.
- Disturbances can be compensated for, which implies that high diagnosis performance can be obtained in spite of the presence of disturbances.
- It is generally applicable to more kinds of components. Not all hardware can be duplicated.
- No extra hardware is needed, which means lower cost, weight and space requirements

The disadvantage of model based diagnosis is the need of a reliable model which calls for good system knowledge and possibly leads to a more complex design procedure.

2.1 Diagnosis system

By using available information about the process we want to extract one fault mode that can explain the process behavior. Sometimes there are more than one fault mode that can explain the process behavior and this should be reflected in the diagnosis statement. To

design a model based diagnosis system, a model of the process is needed. Models of possible faults and how they affect the process are also needed.

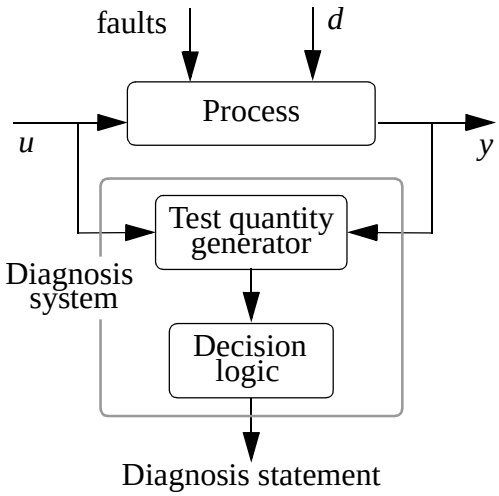


FIGURE 1. Diagnosis system, information flow

In Figure 1, an overview of a diagnosis system is seen. The information available to a diagnosis system consists of measured signals *y*, and control signals *u*. The behavior of these signals are supervised in order to make a diagnosis statement. The unknown signals, *faults* and disturbances *d*, affect the process but can not be measured. The available signals, *u* and *y*, are fed into an algorithm to generate a number of test quantities. This algorithm, the test quantity generator, is based on a model of the process and models of possible faults. The test quantities should be constructed to output a value of zero when no fault is present and non zero when a fault occurs. The test quantities are fed through the decision logic where they are used for taking decisions in a number of hypothesis tests. The decisions are combined to a diagnosis statement.

2.1.1 Model

In the work with model based diagnosis it is important to have a reliable model of the process. The building of a diagnosis model requires good process knowledge, not for achieving a very accurate model but to extract the most important behavior of the process. Even if model accuracy directly affects the maximum diagnosis performance it is shown that rough models can be used with success. An example of this is the rough valve model used in Chapter 2.3.3 and later in the developed diagnosis system in Chapter 4.

2.1.2 Fault models

Not only a model of the process is needed, but also models of all faults the system is supposed to detect. Faults not considered in the design of the diagnosis system will have an unknown affect on performance. To reduce the risk of false alarm it is important to be aware of which faults are the most common. To achieve a well designed diagnosis system, good knowledge of possible faults is required. For a system which is not in production, tools like FTA and FMEA [9] can be used to analyze the system and isolate critical and likely failures. In the Gripen case, an aircraft with many hours in the air, a lot of experi-

ence has been gathered which gives an empirical knowledge of faults important to supervise.

A set of fault modes are defined to explain the state of the process. In each instant, the state of the system is assumed to be one of the predefined fault modes. A natural approach is to assign each fault to one fault mode. A fault mode representing the process in a fault free state is also considered. This approach works fine as long as only one fault occurs at the time. If two or more faults are likely to appear at the same time, they must be assigned to an additional fault mode. If not, a proper diagnosis statement might not be made by the diagnosis system in that case. The idea with fault modes is to name each possible state of the supervised process. The diagnosis system is then supposed to give a statement of which fault mode actually is present. If that is not possible the statement should be a list of possible fault modes.

2.1.3 Hypothesis test

A decision between two possibilities is generally called a hypothesis test. The idea with hypothesis tests in a diagnosis system is to figure out which state the supervised process presently is working in. A decision is made, to tell in which of two sets of fault modes the actual state is found. A test quantity is used for taking the decision.

Test quantity. The purpose with test quantities is to find deviations from normal process behavior. Mathematical relations from the process model are used to describe how signals should relate to each other. Building of test quantities is restricted to available signals from sensors and known inputs to the system.

A designed test quantity is constructed in such a way that it is not affected by all fault modes. Fault modes that do not affect the test are decoupled and the hypothesis test becomes a test between the set of decoupled fault modes R, and the complement R_k^C.

$$H^0: \quad F_p \in R_k = \{\text{set of decoupled fault modes}\}$$
$$H^1: \quad F_p \in R_k^C = \{\text{all other fault modes}\}$$

If H^0 is rejected we assume H^1 holds true. This assumption requires that every possible fault mode is considered. At least all fault modes that the diagnosis system is supposed to handle. Non considered fault modes will have an unknown affect on the diagnosis statement.

Thresholds. Because of disturbances and measurement noise, the test quantity is usually not exactly zero in the fault free case. Therefore, we need to use a threshold with the test quantity to take the decision between the two hypothesis.

$$H_k^0 \text{ is not rejected if } T_k < J_k$$
$$H_k^0 \text{ is rejected if } T_k \geq J_k$$

The convention used here and also common in hypothesis testing literature [10] is to assume that H^1 is true when H^0 is rejected, if H^0 is not rejected we will not assume anything.

Statement. The result from a test is the statement S_k,

$$S_k = \begin{cases} S_k^0 = \Omega & \text{if } H_k^0 \text{ is not rejected} \\ S_k^1 = R_k^C & \text{if } H_k^0 \text{ is rejected} \end{cases}$$

where Ω denotes the set of all fault modes. The statement is a list of possible fault modes that can explain the process behavior. A test quantity with a small value can be explained by all fault modes. If the test quantity exceeds the threshold, the fault mode that can explain the data is not decoupled. The null hypothesis is rejected and the alternative hypothesis is accepted. The alternative hypothesis says that the present fault mode belongs to the set of non-decoupled fault modes.

2.1.4 Isolation

Fault isolation can be performed using several different principles. The approach used here is a structure of hypothesis tests. This makes it possible to diagnose a large variety of different types of faults within the same framework and the same diagnosis system.

A number of hypothesis test are performed individually, each one coming up with a statement S_k. The statement from each test is a list of possible fault modes. The final diagnosis statement S, becomes the intersection of all sets S_k.

$$S = \bigcap_k S_k$$

This implies that the diagnosis statement can contain more than one fault mode. This corresponds well to a desirable functionality to get a list of possible faults when more than one fault mode can explain the process behavior.

2.2 Design of test quantities

A test quantity T_k, should be designed such that if a fault mode in R_k^C can explain the process behavior, then T_k should be large. On the other hand if the data match the hypothesis H^0 i.e a fault mode in R_k can explain the data then T_k should be small. After modeling all faults, each fault mode can be seen as an own model and T_k can be used as a measure of the validity of the different models.

2.2.1 The prediction error principle

Using a model and measurements, it is possible to make a prediction $\hat{y}(t)$ of the output *y(t)*. the prediction error is then a natural measure of the validity of the model. The test quantity can be written as

$$T_k = \min_{\theta_i} V(\theta, x) \qquad i \in \Omega$$

where Ω is the set of all fault modes and θ is a set of parameter values for the model corresponding to fault mode *i*. V(θ,x) is a measure of the validity of the same model. Using the

prediction error principle a sum of prediction errors is used to measure the validity of the model.

$$V(\theta, x) = \frac{1}{N} \sum_{t=1}^{N} \left\| y(t) - \hat{y}(t|\theta) \right\|$$

if θ only can have one value θ_0, the test quantity becomes

$$T_k(x) = V_k(\theta_0, x)$$

and no minimization has to be performed. In most cases to minimization is quite straightforward, but in some cases it may cause heavy computational load. In such a case it is possible to relax V(θ,x) with another function $\overline{V}(\theta,x)$ with optimum close to each other. Then it is reasonable to use $\overline{V}$ in the calculation of the test quantity and expect approximately the same result.

2.2.2 The parameter estimation principle

Using measurements it is possible to estimate parameters in a model of the process. If the estimated parameter has a nominal value θ_0 corresponding to a fault free case, a test quantity can be constructed by directly using the estimated parameter.

$$T(x) = \left\| \hat{\theta}(x) - \theta_0 \right\|$$

Test quantities based on estimates can have very good performance for a fault mode corresponding to the estimated parameter. For other fault modes the performance might be quite low.

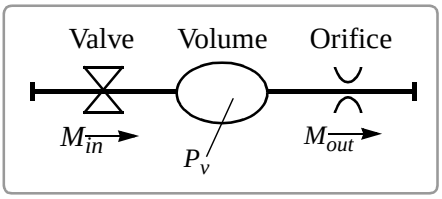
2.2.3 Observers

The building of observers for a diagnosis system are often made according to two common strategies, dedicated observers and generalized observers. Both uses all process control signals and a subset of the measured signals. A dedicated observer is driven by only one measured signal which implies that faults in all other signals are decoupled. The direct opposite situation is used with generalized observers when all but one measured signal are used to feed the observer. Such an observer has only one sensor signal decoupled. These two strategies are used to get a systematic approach for constructing observers, of course combinations of these are also used when the number of used signals are chosen freely.

2.2.4 Example of transient suppression

This is an example of how a model can be used to suppress the effects of dynamics in a process. Two test quantities are developed, one static and one dynamic.

Consider a volume with air flowing in and out of it. The flow is driven by a constant inlet and outlet pressure. A valve is attached to the inlet pipe and the outlet passes through an orifice. The aim is to distinguish between leakage and transients caused by a change in the valves position. Model errors and sensor noise are present.



Static test quantity, T_{stat}. The static larm is based on the assumption that the flow in equals the flow out of the volume. The process model

$$M_{in} = M_{out}$$

is used for constructing a test quantity. Measuring both flow in M_{in}, and out M_{out}, a test quantity could be calculated as

$$T = |M_{in} - M_{out}|$$

The process model does not hold true in a sudden change of valve position. To reduce the affect of noise and model error during transients, the test quantity is low-pass filtered

$$T_{stat} = LP|M_{in} - M_{out}|$$

$$T_{stat} = \frac{1}{s + 0.1}|M_{in} - M_{out}| \tag{2.1}$$

In this test, flow in and out of the volume are measured and used to calculate the test quantity. If an additional signal for the pressure in the volume is available, it can be used to model transients. The next test quantity T_{dyn}, is based on a dynamic model and an additional measure of the pressure P_v.

Dynamic test quantity, T_{dyn}. A more precise model consider the fact that we have a volume between in and outlet flow. An equation based on mass continuity will bring time dependency to the process and the model gets dynamic. The equation of continuity for a fluid

$$\dot{P_v} = \frac{RT}{V}(M_{in} - M_{out})$$

is used as a model of the process. A test quantity is calculated from measure of the two flows and pressure P_v, in the volume

$$T_{dyn} = \left|M_{in} - M_{out} - \frac{V}{RT}\dot{P_v}\right|$$

This test quantity is also low-pass filtered to suppress noise and. The filter uses used in T_{stat} is used once again to achieve the same noise level and a fair comparison of the two test quantities. Now the expression yields

$$T_{dyn} = LP\left|M_{in} - M_{out} - \frac{V}{RT}\dot{P}_v\right|$$

$$T_{dyn} = \frac{1}{s+0.1}\left|M_{in} - M_{out} - \frac{V}{RT}sP_v\right| = \left|\frac{1}{s+0.1}(M_{in} - M_{out}) - \frac{s}{s+0.1}\frac{V}{RT}P_v\right| \quad (2.2)$$

The pressure is considered as a measured signal, volume and temperature as constants.

Simulation. A simulation is performed with the two test quantities working independently at the same time. The test quantities are found in Figure 2, T_{stat} (2.1) is seen to the left and T_{dyn} (2.2) to the right. First a step in the valve position is made at time, t = 2sec, and then a leakage occurs at t = 8sec.

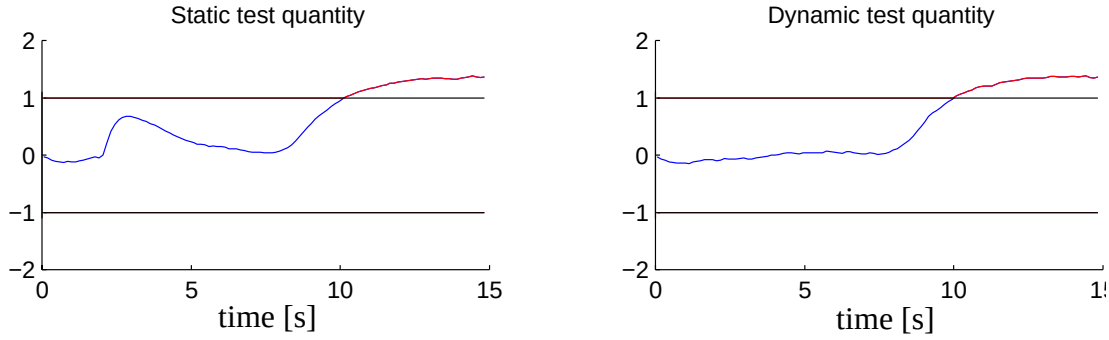


FIGURE 2. System transient followed by leakage

The advantage of using a dynamic model is obvious. The static test quantity almost fire a false alarm during the transient, but the test quantity based on a dynamic model gives a good dynamic behavior. A drawback with the dynamic approach is the extra information needed to make use of the process model. Information about the parameters volume and temperature and a signal for the volume pressure is needed.

In this example, a sensor is considered to deliver the pressure signal. The use of an extra sensor may seem unfair when comparing the two tests. The signal could also be calculated with an observer, based on knowledge of ambient conditions and models for fluid flow in and out of the volume. If the extra sensor were not used, a more fair comparison of the two test quantities would be achieved, but in this way the example is more easy to follow. The purpose with the example is still shown, the benefits with a model based approach for handling transients.

2.3 Thresholds

Thresholds can be chosen in many different ways. In this work two approaches are used. The first, *maximum deflection*, is restrictive and easy to automatize. It is used to give all thresholds an initial value. The second approach, *histogram inspection*, is used for tuning the thresholds to improve diagnosis performance. Both approaches are based on Monte Carlo simulations. No two simulations will give exactly the same result since parameters are randomly distorted and sensor noise is present. A large number of simulations are per-

formed and evaluated, the results will hopefully cover model errors and correspond to what would come from a perfect model.

The adaptive thresholds are of a different nature and can not be chosen in the same way as the other thresholds. The adaptive thresholds are chosen manually from knowledge of process and model errors.

2.3.1 Maximum deflection

Since sensor noise and parameter uncertainties are included in the model no simulations will give the exact same result. After the test quantities are designed, a long range of simulations are performed. Each simulation will give a different result and thresholds are given values larger than the maximum deflection of the corresponding test quantity. The simulations are performed at different working conditions with steps in all control loops to excite the system in different directions.

This approach to find thresholds is straightforward, systematic and easy to automatize but there is a risk of being to restrictive. Some thresholds must be set very high to guarantee no false alarms. The problem with thresholds set too high is that they will never fire, not even when a fault occurs.

A more reasonable approach allows a small risk for false alarm, which will lower the threshold and decrease the risk for missed detection. The test quantity distribution can be estimated from Monte Carlo simulations. To achieve a good estimate of the tail of the distribution a lot of simulations are needed. Approximations of tail distributions from a fewer, more reasonable, amount of simulations is discussed in [12].

After thresholds are given initial values with the *maximum deflection* approach, diagnosis performance may be improved by inspecting histograms.

2.3.2 Histogram inspection

A more fair approach is to set the limit as a compromise between risk of false alarm (threshold too low) and risk of not detecting a fault (high threshold). A correct built test quantity is zero for decoupled fault modes and high for others. In Figure 3, two distributions are seen, showing the typical behavior of a test quantity. One for a fault free system and the other for a system with a non-decoupled fault present.

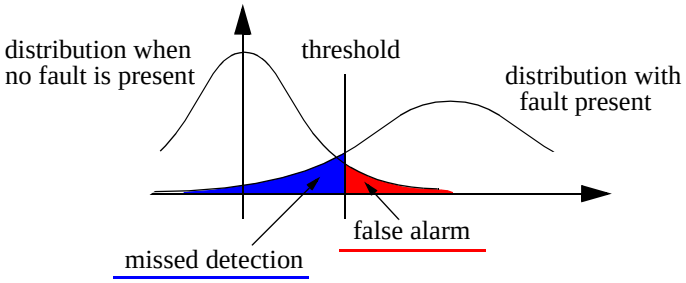


FIGURE 3. Test quantity distributions

The left distribution shows the test quantity when no fault is present or with the present fault decoupled. The test quantity is distributed around a value of zero. The tail of the distribution found higher than the threshold corresponds to the risk of false alarm.

The distribution to the right in the figure, shows the typical behavior when a fault affecting the test quantity is present. The present fault mode is found in the alternative hypothesis for the corresponding test. The tail of the distribution, lower then the threshold corresponds to the risk of not detecting a fault.

The idea with histogram inspection is to set the threshold by considering the two requirements of low false alarm rate and low risk of missed detection. It might be worth lowering the threshold to increase the possibility to detect faults, on cost of increased false alarm rate.

In Figure 19 on page 41 an example is seen. A test quantity from the diagnosis system designed in the work, is evaluated. The histogram comes from 12 simulations with different fault modes present. By lowering the threshold, the missed detection probability is decreased and diagnosis performance will improve.

2.3.3 Adaptive thresholds

The concept of adaptive thresholds is based on knowledge of model errors. To illustrate this, a valve is chosen as object for supervision. The dynamics of the valve is known and approximated with a model of reduced order. The approximation will introduce model errors with known characteristics. An expression of the model error is used to obtain the adaptive thresholds.

Valve model. A low order model of a valve is studied. The model is used to predict the valves position and a test quantity is formed according to the estimation error principle. Assuming that no faults are present, every valve will reach its settling point within a certain time. In steady state it will always be in the correct position. In other words, at low frequencies the valve settles correctly, which motivates a model with the transfer function

$$G(s) = \frac{1}{0.1s + 1} \tag{2.3}$$

The valve model transfer function *G(s)*, is shown in Figure 4 below.

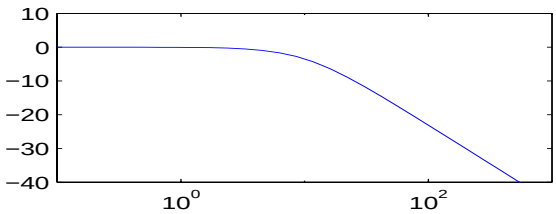


FIGURE 4. Valve, low order model transfer function, G(s).

Model error. When the set point changes fast, the position is harder to predict. The model is not reliable at high frequencies, the higher frequency, the higher the model error

becomes. We can assume a model error that is low at steady state and large when the valve is moving fast. The transfer function, ΔG(s) is used to tell the size of the model error.

$$\Delta G(s) = \frac{0{,}1s}{0{,}1s + 1} \tag{2.4}$$

The model error transfer function $\Delta G(s)$, is shown in figure Figure 5.

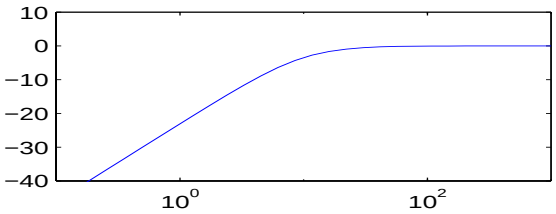


FIGURE 5. Model error transfer function, ΔG(s).

A real test. The assumptions above have been tested in a test quantity for evaluating data from a real valve in the Environmental Control System. The low order model (2.3) of the valve is used to predict the position. A test quantity is built with the prediction error principle according to

$$T = |G(s)u - y|$$

This test quantity is used together with the adaptive threshold, J_{adp}. The threshold is derived from Equation (2.4) and yields

$$J_{adp}(t) = k(|\Delta G(s)|u + c)$$

The test works fine most of the time but not in the case of large steps in the position setpoint. This comes from the analog controller that governs the valve, see Figure 6. The voltage that runs DC-motor in the servo is limited and makes the valve to a non-linear system with limited bandwidth. To overcome this problem the position setpoint is limited to a maximum step size before estimating the position. The limitation makes the estimation sensitive to the chosen maximum step value. This extra uncertainty should be considered in the calculation of the threshold.

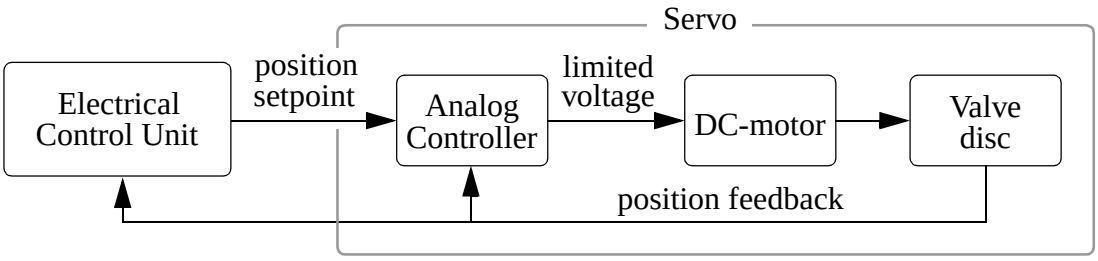


FIGURE 6. Valve overview

A test obtained from the test quantity together with the adaptive threshold above, is used to evaluate data measured on a real valve. Later in Section 4.1.5 the same test quantity is

used for evaluating simulations on the principle model. Below, two figures are shown for evaluation of measured real data. In Figure 7 the measured position is seen within an estimated allowed range. In Figure 8 the test quantity is seen together with the corresponding adaptive threshold.

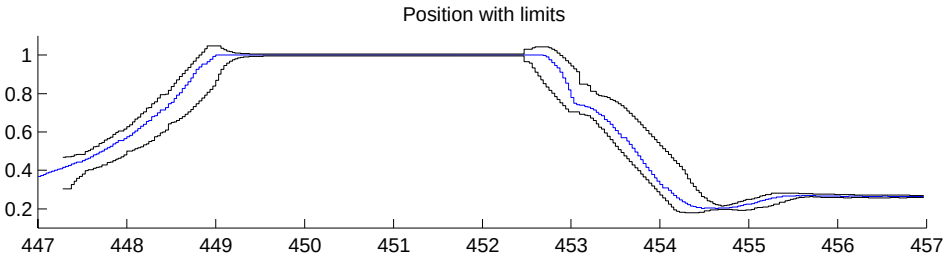


FIGURE 7. Measured position within limits

The measured position is plotted and compared to an estimated allowed position range, derived from the position setpoint. The allowed range *y*, is calculated as

$$G(s)u - |\Delta G(s)u| \leq y \leq G(s)u + |\Delta G(s)u|$$

No data for a faulty valve was available, instead simulations are performed to see if the test can detect valve failures. In Section 4.4 the same test quantity is used for evaluating simulations with a valve model. The test is shown to generate alarms when a fault occurs.

The idea with adaptive thresholds is shown in Figure 8. The test quantity is plotted together with the adaptive threshold. The evaluation is performed on the same data as in Figure 7.

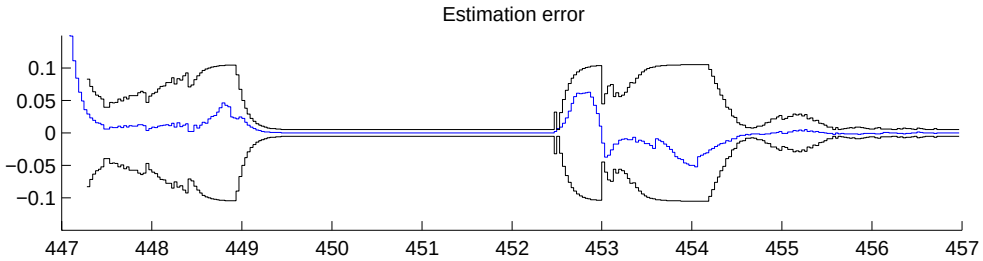


FIGURE 8. Estimation error with adaptive threshold

The estimation is very accurate during static conditions but because of the low model order it can not obtain perfect transient behavior. It is seen how the adaptive threshold increases the allowed region during transients and that it is needed to not get a false larm. If adaptive thresholds were not used, the threshold had to be set to the maximum value in the simulation and the performance of the test would be lower.

2.4 Requirements

As in all system design, is it important with requirements to get a measure of the system performance. When designing a diagnosis system, the three most common requirements are, *false alarm*, *missed detection*, and *time delay.*

False alarm rate. Is the risk of getting a larm even when there is no fault present in the system. The reliability of the system is directly connected to this requirement. An operator that feels that the system has a high false alarm rate, might not take alarms serious and even ignore them. The *false alarm rate* increases with lower thresholds. With low thresholds, alarms fire more easily.

Missed detection. The risk of not detecting a fault present in the system. When designing a restrictive diagnosis system, that does not fire alarms easily, the risk increases for not detecting an alarm. The probability for *missed detection* increases with higher thresholds, therefore this requirement in many cases goes in conflict with the requirement of low *false alarm rate*.

Time delay. Requirements for the time it takes for the diagnosis system to detect a fault are important. The most common requirement is the *mean delay time*, the mean time it takes to detect a certain fault mode. Another common requirement is the *maximum delay time*, an upper limit of how much time that can pass by before the diagnosis system detects a fault.

3 The Gripen Environmental Control System

This work is performed at the department of General Systems which is responsible for several systems as hydraulic, electrical power, fuel and environmental systems. In order to work with model based diagnosis one of the systems was chosen to focus on. The choice fell on the Environmental Control System, ECS. The ECS is a system with many interesting characteristics, such as being dynamic, non-linear and large. The system is built up from numerous components and the physics behind compressible fluids gives a highly nonlinear system with dynamic behavior.

Redundancy management is built into the system to guarantee performance. The redundancy consists of auxiliary units and possibility to shut off parts off the system. Supervision is required to manage this system redundancy. It is desirable to increase reliability and functionality of this supervision, and to extend it with a diagnosis system to automatize fault localization.

3.1 Environmental Control System, ECS

The Environmental Control System, ECS, has a number of different tasks to perform. The most important are pressurization of cabin and cooling of electronics. It is important to always keep a pressure in the cabin that does not distress the human body. The cooling of some electronics such as the flight computer is also of great importance since it is necessary to keep the fighter in the air.

Other, less critical tasks also has to be performed, like comfort to the cabin, defrosting the windshield and pressurization of tanks and gearboxes. An onboard oxygen generator will eventually be installed and also driven by the ECS.

The ECS can schematically be divided into three parts: *Air Supply*, *Air Conditioning* and *Distribution*. The three parts can be found in a schematic overview of the ECS, in Figure 9.

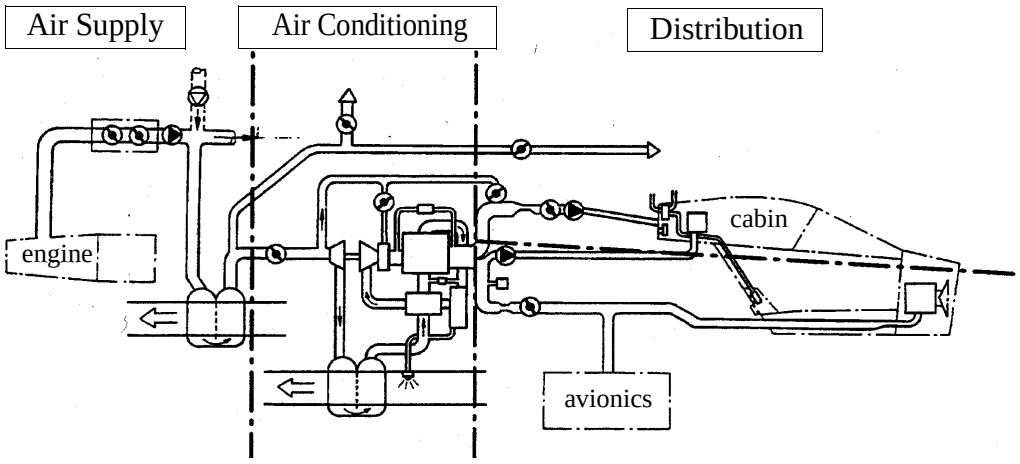


FIGURE 9. ECS overview

The ECS system is supplied with air from the engine, so called bleed air. The bleed air is hot and at high pressure that must be reduced to more handy working levels. The pressure

is reduced in a pressure reducing valve and the temperature is decreased in the primary heat exchanger. A bypass valve, around the heat exchanger, is used to get a first coarse control of the temperature. After this first step of adjusting the air temperature and pressure, a fraction is distributed to the defroster and for pressurization of tanks and gearboxes.

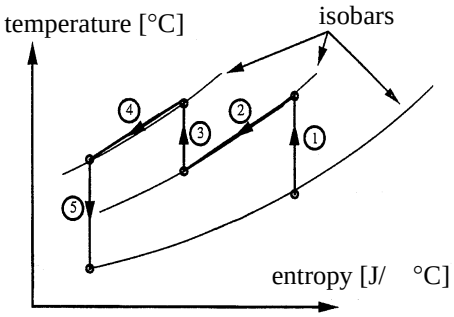
When the engine is off or if it is not desirable to load it more than necessary, additional air supply can be provided by the use of an Auxiliary Power Unit, APU. The APU is also used to deliver air for starting up the engine.

The next section, the air conditioner, which is also called the cooling pack, starts with a valve used to control the pressure at the outlet of the cooling pack. The next step in the process of air conditioning is to send it through the cold air unit. The pressure is increased in a compressor which results in increased temperature. The air passes through the secondary heat exchanger followed by the condenser and an efficient water separator. Dry air is of great importance to sustain a long lifetime of the electronics.

The last step in the cooling process is to expand the gas through the cooling turbine. Now the air has a temperature below the freezing point and in order to reach the desired temperature of 0°C, hot air is bypassed the cold air unit and mixed with the outlet of the cooling pack.

The distribution part has two main pipes, one for the cabin and one for the electronics. The temperature to the cabin is once again mixed with hot air to get the comfort in the cabin the pilot wishes. The flow is controlled separately to each branch.

The idea with the cooling pack is to remove moisture from the air, and to lower the temperature below ambient temperature. The course of events can, from a thermophysical point of view, be divided into five steps.



1. Compression of ambient air in the engines high pressure compressor.
2. Cooling in the primary heat exchanger. Heat is transferred to the ambient air.
3. Compression in the cold air unit compressor. Compression of the air is needed to easily achieve an additional heat transfer to the ambient air.
4. Cooling in the secondary heat exchanger. Removing more heat from the air. Now, the energy in the air is low enough to get the required temperature.
5. Expansion in the turbine of the cold air unit, with additional temperature loss as a consequence.

Two conditions must be fulfilled to get a proper cooling. The first condition is the need of bleed air supply, quite obvious. Some extreme operating points may violate this condition, but is easily overcome by increased engine thrust. The second is the need of external air passing the primary and the secondary heat exchanger. The air flow depends on aircraft velocity. On ground and at low velocities the flow can be increased by the use of ejectors that take high pressured air from the ECS system, and uses it to accelerate the flow through the ram air channel.

Cabin pressurization above external ambient pressure is required to maintain a breathable atmosphere and avoid human distress at high altitude. In civil aircraft this results in cabin pressure being maintained at a value equivalent to altitudes no greater than 2400 metres and a requirement that cabin altitude should never exceed 5500 metres, not even in an emergency. Passenger comfort also dictates limitations on the rate of change of pressure. For military aircraft similar considerations apply but, since the crew are deemed to be fit and healthy and breathing oxygen is constantly supplied, the requirements are considerably relaxed. Cabin pressure differential is reduced to minimum in order to minimize structural weight.

3.2 Opinions and reflections about the ECS

A lot of information about the ECS and about diagnosis of the system has been gathered. This involved both reading of files from archives and talking to people involved in these issues. Here is a collection of thoughts and reflections encountered during the work.

The general opinion about the ECS is that so far it has been a quite problematic system. This is mainly due to outsourcing of system development and production to a subcontractor. Supervision and diagnosis was added at a late stage in the development process and as a a consequence a lower level of accuracy in the requirements was held.

Initially the diagnosis system suffered from many false alarms and undetected faults. Since the system was not developed at Saab there was no easy way to modify it. Some problems could be solved by adjusting thresholds and installing extra sensors, whereas some alarms simply were shut off.

The main problem with the diagnosis of the ECS is not to detect faults, but to isolate them. As the diagnosis is implemented today, it mainly supervises system output to be close to the setpoint. This sometimes cause the diagnosis system to fire an alarm when the ECS is saturated. Not entirely perfect, since an alarm is not only supposed to give a warning, but also a recommendation of suitable measures to take. An alarm caused by system saturation has a completely different message to the pilot than a larm caused by system failure. Another advantage with better isolation capabilities in the diagnosis system would be to facilitate maintenance on the ground. Today isolation are performed manually, a work that would be made a lot faster with a well working diagnosis system.

By experience it is well known what the most common faults are. It seems like the mechanic construction is very robust and problems occur almost only in parts with electric connections, like valves and sensors. The valves life is shorter than expected, probably due to non tuned control laws. A worn out valve shows symptoms like stiction with discontinuous position changes. Other symptoms are slower movements and increased friction. Other faults known to occur on the valves are displaced potentiometers, displaced valve discs, and loose connections. Displaced potentiometers results in a biased feedback and the valves settles at the wrong position. Displaced valve discs leads to an incomplete operating range. Sensor connections that are loose give a total loss of signal every now and then.

Another thing worth to notice is the lack of redundancy. Once again, not from a mechanical point of view. Crucial tasks are covered with auxiliary systems. Also other safety arrangements like safety valves appear. However, as mentioned earlier, the electrical part has some drawbacks. Each control loop operates only with one sensor and one actuator,

even when redundant components exist. This makes the system sensitive to faults in such a component.

One reason that the supervision is not working satisfactorily is that the ECS have been modified in several steps since the start, but not the supervision. Old thresholds and other limits should be updated to fit the system of today.

One opinion is to only supervise periods of stable system performance avoiding transient events. This would be easier in some ways, like better understanding of system behavior. An existing model is verified and usable for simulation. This approach would still leave some questions. How do you decide when the flight is static or not? Is there enough statical time in a typical flight session?

From an economic point of view is it not obvious to automate all diagnosis. Tests that must not be run too frequently, may be cheaper to execute manually. An automatic system involves a large cost in development and implementation. On the other hand is usage very cheap. The situation is the direct opposite with a manual procedure that involves less development, but takes a lot of time at each occasion and becomes costly in the long run.

3.3 Diagnosis today

The ECS, manufactured by a subcontractor, has the same supervision and control system as from the beginning. Today it is in a need of revision. Rapid controller loops wear unnecessarily on valves. Supervision does not match the modified system of today and hardly no diagnosis is present. With that in mind Saab, decided to take over the responsibility of the system and develop their own controller box.

The supervision of the ECS as it looks today can be divided into two categories, as the terminology at Saab goes; *Fault Monitoring* (FM) and *Safety Check* (SC).

Fault Monitoring. The Fault Monitoring is run continuously during flight. It is supposed to detect faults in subsystems and when necessary switch to available auxiliary systems. The supervision is a passive on-line system assigned to guarantee safety and performance demands. Today it mainly checks that important quantities like different temperatures and pressures are within specified limits. The most important of these are cooling capacity to avionics and pressurization of the cabin.

Safety Check. Safety Check is performed at startup before each flight. It is done once and must be quite fast without leaving uncertainty about system safety. The test is truly automatic and runs before, during and a short time after engine start. If the test is run clear, the aircraft is supposed to be safe and can be taken in the air. The test is also run at service occasions. The tests performed here are online and combined active and passive. Communication between computers is one big test, another is fast repositioning of the valves to check the settling times.

Actually, there is one more category, called Functionality Check (FC) but it is not an automatic system and not considered in this work. The Functionality Check is a length, time demanding procedure, performed in cooperation with an operator only at special service occasions. Most steps are actually completely manual tests that checks performance and safety. Inspection of installations, performance of controllers, and functionality tests of redundant systems that can not be performed automatically. Examples are inspection of

leakage in pipes and functionality of safety valves. The most important task is the test of redundant systems which are not needed in normal flight, but critical for safety at failure of other parts.

The procedures mentioned above are implemented with the intention to guarantee safety and performance. A focus is held on detecting deviations in system output and finding faults on a system level. For maintenance purposes it would be desirable to isolate faults, to track them down to a component level. When detecting a fault, the supervision gives a recommendation of suitable measures to take. Only in a few cases this leads to isolation of likely faulty components. Instead, the most common recommendation is MFL, Manual Fault Localization. This has given the service crew a lot of experience in manual diagnosis of the system. A tool for propagating this knowledge is developed, which here is referred to as symptomatic diagnosis. As the name reflects, it is based on examination of deviations in the general behavior of the system.

To meet customer demands a maintenance manual [6] has been written. Normally such manuals are based on Fault Tree Analysis, FTA [9], but this is not suitable in the ECS due to its size and complexity. A complete fault tree analysis would fast become very large and unwieldy. Instead a *symptom based diagnosis* procedure is inserted in the manual.

Symptom based diagnosis. With the symptom based diagnosis procedure, the fault isolation is performed off-line, manually, by taking a closer look at recorded signals. The idea is to extract interesting behavior according to predefined criteria and match in a fault matrix to generate a list of probable faults. The procedure can be divided into four steps.

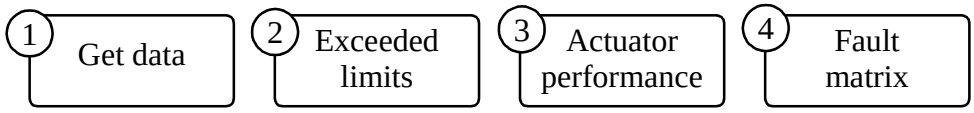


FIGURE 10. Procedure for symptom based diagnosis

1. Examine registered data from the actual flight. Extract a region without transients close to the event of the alarm.
2. Find sensor signals that have exceeded their limits and make a mark if the signal is too high or low.
3. Look at the corresponding actuator performances. Every quantity has one actuator to control it. If the actuator quantity has exceeded a tolerance limit, the actuator should be saturated. Mark a yes for correct saturation, and no otherwise.
4. Match the combination of high/low/yes/no in a matrix to identify a fault or a list of possible faults. The matrix consists of a row for each fault and a column for each symptom. By matching a combination of symptoms in the matrix, a list of possible faults can be found.

Symptom based diagnosis has been tested for some time in the swedish airforce, with positive feedback from those who have tried it. A thought that immediately occurs is that it seems to be a very systematic procedure, well suited for automation, and possible to implement as an automatic built in test. The problem is to know when signals have exceeded their tolerance limits. No firm limits exist and under some circumstances limits can be stretched, e.g. at system transients.

3.4 Motivation for the principle model

In order to perform model based diagnosis, a model of the process is needed. A model of the ECS already exists, built in a simulation software called Easy5. This model exhibits good performance in static simulations, but has not been verified in dynamic cases. Since this work involves dynamic aspects, the existing Easy5 model is not suitable. Instead a principle model is built in Matlab/Simulink, Appendix B: The principle model in Simulink. A lot of effort was made to convert the model from a component-level to a functionality-level. Not every single component has been modeled, but the most important functionality is present. The intention is to extract behavior from ECS to get a base for exemplification of diagnosis concepts.

The principle model embraces the main features of the real system. In Figure 11, the three parts of the ECS are found, Air Supply, Air Conditioning and Distribution. A valve for each control loop is also seen in the figure. Distribution to the two main air-consumers, avionics and cabin, are present. Two control loops for the flow distributed, one for each consumer. An air conditioning part with control loops for temperature and pressure. Air supply with hot, highly pressurized air. The aim was first set to handle all five main control loops, one to control cabin temperature in addition to those mentioned above. After some considerations the control of cabin temperature was removed. As well as the bypass pipe to perform this functionality. This reduction of model order made the simulations run faster.

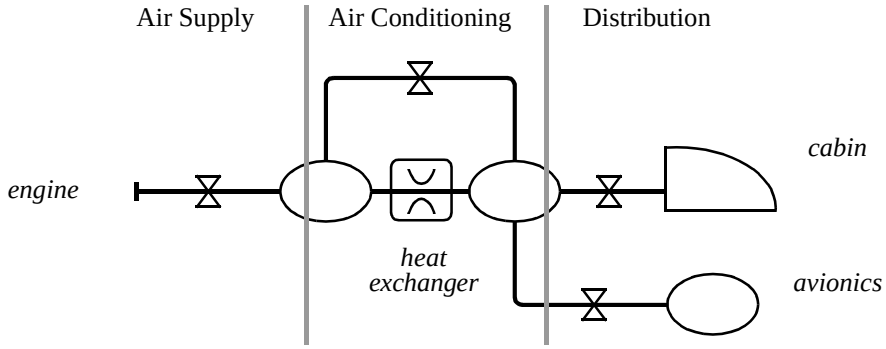


FIGURE 11. Principle model with main features

The model starts with a valve, found in the real system downstream the primary heat exchanger. This means that the air actually already is given a coarse regulation before it enters the model.

The cold air unit, that in reality consists of components like compressor, turbine and two heat exchangers, is here present as only one heat exchanger. This is motivated with the major purpose of the unit, cooling of air.

The air-conditioning part consists of a heat exchanger with a bypass valve to control outlet temperature. This construction symbolizes the cooling pack. A temperature sensor at the outlet of the cooling pack is used to control the valve. The two distribution pipes are connected at the outlet.

As mentioned earlier, all cold air users are not included in the model, e.g. defrosting of the windshield and pressurization of tanks and gearboxes. Consumers not needed all the time, like radar and ejectors, are not modeled, or hypothetically shut off.

Consecutive pipes and compartments have been added together to one volume. Distribution pipes with the same origin and destination have been replaced with one approximated orifice.

Parameters in the principle model have been extracted from the model in Easy5 as far as possible, Appendix A: Model parameters.

The purpose has been to extract the typical behavior of the system, not to keep track of every fraction of flow distributed to the more than fifteen electronic boxes.

3.5 Air components

In order to build a model of the ECS, based on physical relations, knowledge of components and air dynamics is needed. In this chapter, all components needed for building the principle model are described. The principle model is developed in Chapter 3.6.

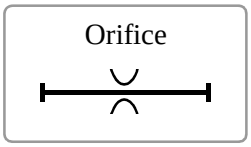
Building of models in simulink is made a lot easier if the system is divided into different objects. In this chapter components needed for building the process model are described. Flow and pressure are the two primary quantities needed for simulation of the fluid dynamics. These are calculated from relations for flow restrictions and volumes. Flow restrictions are modeled with the two components orifices and valves. A component for describing the heat exchanger is also used.

3.5.1 A Nozzle

A flow restriction with minimal energy loss would be a nozzle. A nozzle has a well rounded inlet and a smooth outlet profile. The nozzle have some peculiarities like it chokes when the flow stream velocity reaches a sonic limit. Beyond the sonic limit no further pressure difference can increase the stream velocity. Further such a device can, with minimal energy loss, be used for indirectly measuring the air mass flow. The air flow calculated from measure of static pressure loss in the restriction and temperature and absolute pressure upstreams the restriction. Such a arrangement is better known as a venturimeter and used in the ECS for measuring and controller feedback of flow distribution to cabin and avionics.

The nozzle is not used in the principle model, instead orifices are used to model flow restrictions. The orifice is a more correct model when dealing with turbulent fluids.

3.5.2 The Orifice



The classic orifice is characterized by a circular, sharp edged hole mounted in a circular duct. Since the flow enters the orifice without favorable inlet shaping, the cross section will continue to contract for a short distance downstream. The mass flow rate will therefore be less than for a nozzle having the same throat diameter. More details about flow restrictions can be found in [7].

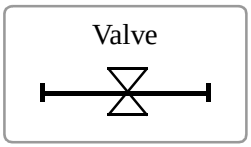
A compressible gas flowing through a sharp edged orifice may be calculated according to the empirical formula.

$$M = \frac{K_0 A}{\sqrt{T}} \sqrt{P_u^2 - P_d^2} \tag{3.1}$$

M = mass flow rate [kg/s]
A = orifice area [mm^2]
P_u = pressure upstream [kPa] abs
P_d = pressure downstream [kPa] abs
T = inlet temperature [K]
K_0 = constant of proportionality = $3,569 \cdot 10^{-5}$

It will be noted that an orifice, unlike the nozzle, never chokes. With an orifice the conditions downstream always affect the conditions upstreams, no matter of sonic limit.

3.5.3 The Valve



Valves are modeled as orifices with variable area. The model consists of two parts, one with the mechanical dynamics for calculating the valves position. The second part with equations for the fluid relations. The input signal to the valve is an angle setpoint which gives a simulated actual angle. The actual angle is transformed to an open area with the relation

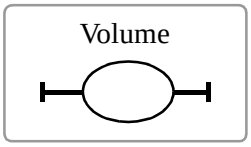
$$A_e = A_0(1 - \cos(a))$$

where

a = valve angle
A_0 = maximum effective open area
A_e = actual effective open area.

The actual effective open area, A_e is used in the fluid relations as a flow restriction with the same characteristics as an orifice.

3.5.4 A Volume



The pressure in a constant volume is calculated by considering the flow in and out of it. The equation of state for a perfect gas, yields

$$PV = mRT$$

P = absolute pressure [Pa]
V = volume [m^3]
m = mass in volume [kg]
R = gas constant = 287 [J/kg/K]
T = temperature [K]

Assuming constant temperature in the volume, and differentiating the equation for state of a perfect gas gives the expression

$$\frac{d}{dt}(PV = mRT)$$

$$\dot{P}V = \dot{m}RT = MRT$$

Assuming mass continuity, the air flowing in and out of the volume must equal the mass change in the volume, $M = M_{in} - M_{out}$. The expression becomes

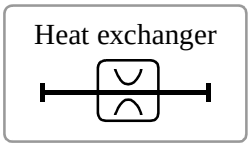
$$\dot{P} = \frac{RT}{V}(M_{in} - M_{out}) \tag{3.2}$$

M = mass change in volume
M_{in} = mass flow in to volume [kg/s]
M_{out} = mass flow out of volume [kg/s]

The expression (3.2) is used for calculating the pressures in the principle model. Each pressure becomes a state in the state-space model explaining the fluid dynamics.

3.5.5 Heat exchanger

The idea behind the heat exchanger is to let two fluids be in thermal contact without mixing with each other. Some different design approaches exist. In Gripen a cross-flow heat exchanger is used, a choice that implies a reduced weight and size.



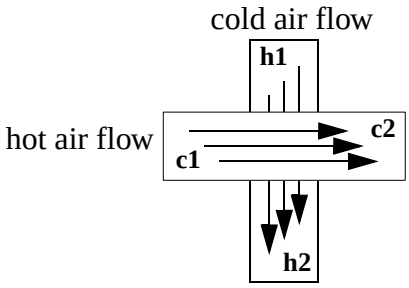


FIGURE 12. Cross sectional heat exchanger

Heat is transferred from the hot to the cold air. Energy balance requires that the energy q, received by one fluid must be emitted by the other fluid.

$$q = M_h C_{p,h}(T_{h1} - T_{h2}) = M_c C_{p,c}(T_{c2} - T_{c1})$$

M = mass flow [kg/s]
T = temperature [K]
C_p = specific heat = 1,006 [kJ/kg/C], (air at 27°C)
indices *h* for hot air, *c* cooling air, *1* for inlet, *2* outlet

When calculating inlet or outlet temperature in a heat exchanger the most common method [1] is based on the effectiveness of the heat transfer. The heat-exchanger effectiveness is defined as

$$\text{Effectiveness, } \varepsilon = \frac{\text{actual heat transfer, } q}{\text{maximum possible heat transfer, } q_m}$$

With knowledge about the capacity of the heat exchanger and working conditions, the temperature loss in the hot air can be calculated by

$$T_{h2} = T_{h1} - \frac{q}{MC_p} \tag{3.3}$$

This expression is used in the principle model for calculating temperature loss in the heat exchanger.

3.6 The principle model

The principle is supposed to simulate the most the typical behavior of the ECS. It also simulates possible faults that may occur on the ECS. Signals corresponding to the sensor signals in the real ECS are extracted from the principle model and fed to a diagnosis system. The diagnosis system is developed in Section 4.

The principle model is composed of a couple of basic components. These are volumes, valves, orifices and one heat exchanger, all found in Section 3.5. The components are combined to a system that reflects the functionality of the ECS. The three steps with Air Supply, Conditioner and Distribution are included. Distribution is limited to the two main consumers, avionics and cabin. The model is not intended to simulate all behavior of the real ECS, but to reflect the most important features and typical characteristics. The work is focused on diagnosis of this system, not getting accuracy with the real ECS.

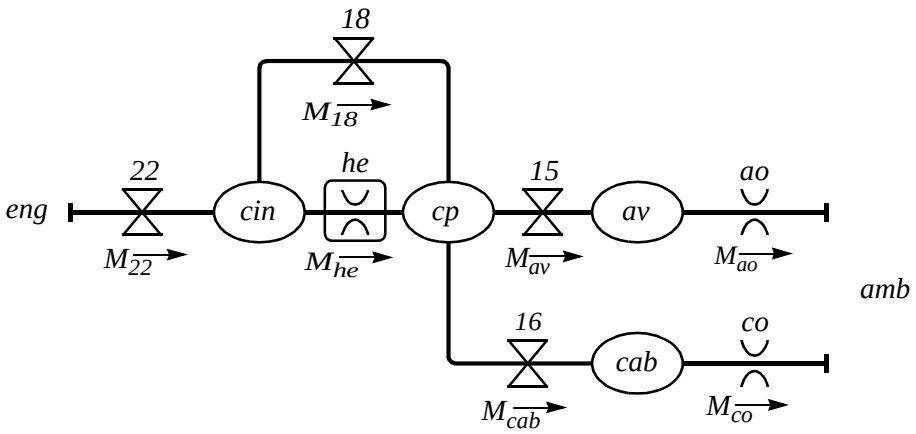


FIGURE 13. Principle model, notation

To keep the size of the model as small as possible some assumptions and simplifications had to be included. The notation is explained together with the model in Section 3.6.1 to Section 3.6.5.

The principle model is described on state space form. In order to get a better overview, the model is divided into two parts, see Figure 14. The first part with calculations of fluid dynamics and the second with calculations of the valve dynamics. The split is suitable

since the valve dynamics are assumed to be determined by valve setpoint signal only, the fluids have no direct influence on valve position.

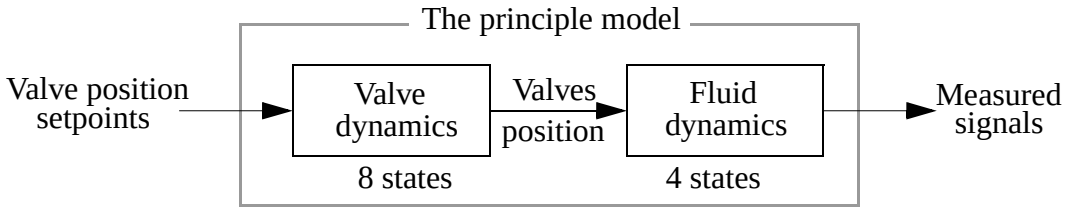


FIGURE 14. Principle model on state space form

The model of the fluid dynamics consists of four states. The valve dynamics consists of four valves, each valve model adding two states to the model.

Fluid dynamics. The fluid dynamics consists of four states, one for the pressure in each volume. The fluid dynamics expressed in state-space form yields

$$\begin{matrix}\dot{x} = f(x,u) \\ y = h(x)\end{matrix}, x = \begin{bmatrix} P_{cin} \\ P_{cp} \\ P_{av} \\ P_{cab} \end{bmatrix}, u = \begin{bmatrix} A_{22} \\ A_{18} \\ A_{15} \\ A_{16} \end{bmatrix}, y = \begin{bmatrix} y_{Pcp} \\ y_{Tcp} \\ y_{Mav} \\ y_{Tav} \\ y_{Pcab} \\ y_{Mcab} \\ y_{Tcab} \\ y_{\varphi 15} \\ y_{\varphi 16} \\ y_{Pav} \end{bmatrix} = \begin{bmatrix} P_{cp} \\ T_{cp} \\ M_{av} \\ T_{av} \\ P_{cab} \\ M_{cab} \\ T_{cab} \\ a_{15} \\ a_{16} \\ P_{av} \end{bmatrix}$$

Valve dynamics.

The state space model of the valve explains the relation between valve setpoint and valve position. This requires two states, position a, and velocity ω. The valve equations are discussed in Section 3.6.4, on state space form they yield

$$\dot{x} = \begin{pmatrix} \dot{a} \\ \dot{\omega} \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -10 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, y = a$$

The input signal u, is the voltage driving the valve dc-motor. The valve model is explained in more detail in Section 3.6.4. Equations for state space models of fluid dynamics are described in Section 3.6.1 to Section 3.6.3. In Section 3.7, the principle model will be extended with model errors and in chapter 3.8 with models of possible faults.

3.6.1 Avionics

The avionics in the ECS consists of a number of pipes for cooling of more than fifteen electronic boxes. All consecutive pipes are replaced with one volume and all outlets are replaced with one orifice with the same total open area. The only outlet in the model leads to ambient, a simplification not totally true since some of the pipes are used for cooling of cabin avionics and therefore should discharge in the cabin compartment. Valve 15 is mounted at the inlet and controls the air flow through the avionics. Equations for the flows to and from the avionics are used to yield an expression for the avionics pressure P_{av}. These are expressed as

$$M_{av} = \frac{A_{15}K_0}{\sqrt{T_{cp}}}\sqrt{P_{cp}^2 - P_{av}^2}$$

$$M_{ao} = \frac{A_{ao}K_0}{\sqrt{T_{av}}}\sqrt{P_{av}^2 - P_{amb}^2}$$

$$\dot{P}_{av} = \frac{RT_{av}}{V_{av}}(M_{av} - M_{ao}) \tag{3.4}$$

3.6.2 Cabin

The cabin is, just like the avionics, modeled with a volume with one inlet and one outlet pipe. The outlet is a constant orifice and valve 16 is mounted in the inlet pipe to control cabin pressure. The equations looks like

$$M_{cab} = \frac{A_{16}K_0}{\sqrt{T_{cp}}}\sqrt{P_{cp}^2 - P_{cab}^2}$$

$$M_{co} = \frac{A_{co}K_0}{\sqrt{T_{cab}}}\sqrt{P_{cab}^2 - P_{amb}^2}$$

$$\dot{P}_{cab} = \frac{RT_{cab}}{V_{cab}}(M_{cab} - M_{co}) \tag{3.5}$$

The model of the outlet as an orifice is an approximation of a mechanical pressure regulating valve. The approximation can be seen as a flow restriction with known but uncertain open area. In Section 3.7.1 all parameters are randomly distorted to give a realistic uncertainty to the model. The open area parameter A_{co}, is modeled with greater standard deviation to compensate the approximation of the regulating valve as an orifice.

3.6.3 Cooling pack

The purpose with the cooling pack is to cool and demoisture the air. Since air moisture is not considered in the model, the cooling becomes the main task of the modeled cooling pack. This motivates the cooling pack to be modeled as a heat exchanger. The heat

exchanger is connected to two volumes, one volume situated at the inlet of the heat exchanger and the other at the outlet. The inlet volume is recognized by the subscript *cin*, and the outlet volume by *cp*. For an overview of the modeled cooling pack see Figure 15.

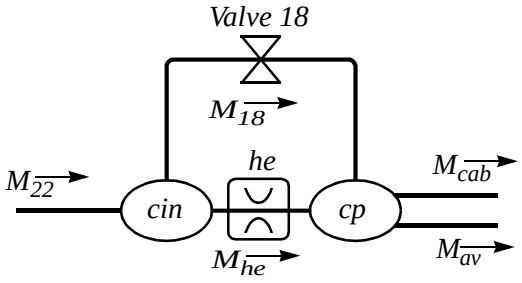


FIGURE 15. Model of cooling pack

In order to get an expression for the states of the air physics expressions for all flows in Figure 15 are needed. The flow M_{22}, in to the first volume, is calculated by

$$M_{22} = \frac{A_{22}K_0}{\sqrt{T_{eng}}}\sqrt{P_{eng}^2 - P_{cin}^2}$$

Two flows are going out of the first volume and in to the second, air flow M_{HE} goes through the heat exchanger and flow M_{18} goes through a bypass pipe. The bypass flow is controlled with vale 18 to achieve a desired temperature at the outlet. The two flows are calculated as

$$M_{he} = \frac{A_{he}K_0}{\sqrt{T_{he}}}\sqrt{P_{cin}^2 - P_{cp}^2}$$

$$M_{18} = \frac{A_{18}K_0}{\sqrt{T_{cin}}}\sqrt{P_{cin}^2 - P_{cp}^2}$$

for heat exchanger and bypass flow respectively. This yields an expression for the pressure at the inlet volume P_{cin}, which expressed on state space form yields

$$\dot{P}_{cin} = \frac{RT_{eng}}{V_{cin}}(M_{22} - M_{he} - M_{18}) \tag{3.6}$$

The flows out of the inlet volume goes straight to the outlet volume. The outlet volume has two flows going out of it, M_{cab} for distribution to cabin and M_{av} for distribution to the avionics. The flows M_{av} and M_{cab} are described in Section 3.6.1 and Section 3.6.2 respectively. Now, the state space equation for the outlet volume pressure P_{cp} can be expressed as

$$\dot{P}_{cp} = \frac{RT_{cp}}{V_{cp}}(M_{he} + M_{18} - M_{av} - M_{cab}) \tag{3.7}$$

The flow through the heat exchanger is cooled down by energy transfer to ambient air. The temperature loss is modeled according to the theory in Section 3.5.5. The expression for the temperature loss becomes

$$T_{he} = T_{cin} - \frac{q}{M_{he}C_p}$$

with the actual heat transfer q, as a function of altitude and velocity, here considered as a constant representing a specific flight case. The air mixture in the outlet volume is considered to be well mixed and reach a mean temperature calculated by

$$T_{cp} = \frac{T_{he}M_{he} + T_{cin}M_{18}}{M_{he} + M_{18}}$$

This equation is based on energy conservation, the total energy in the fluids flowing in to the volume equals to energy in the mixed fluid.

3.6.4 Valves

A general model is used for all valves. Important with the model is to explain the nonlinear behavior of the valve. Most of the nonlinearity comes from limitations in the output voltage of an internal controller. The internal controller governs a DC-motor that actuates the valve position. The valve position is used to calculate the flow resistance through the valve. The model consists of three parts, an internal feedback controller, a mechanical servo motor and a flow restriction. An overview of the valve model is found in Figure 16.

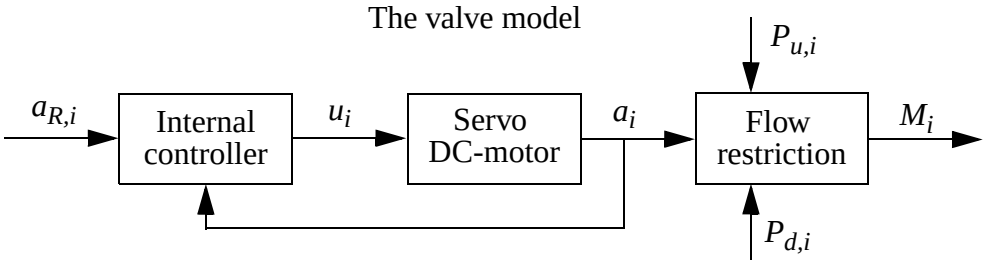


FIGURE 16. Valve model

In order to model the behavior of the DC-motor, a state space model of second order is needed. The other parts does not contribute with any additional states and the total valve model can be expressed as a state space model with two states. Each valve now contributes with two states to the principle model.

The model of the electrical servo is found in [3], parameters are modified to fit measured data from valve 22. The electrical servo model expressed on state space form yields

$$\dot{x}_i = \begin{pmatrix} \dot{a}_i \\ \omega_i \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -10 \end{bmatrix} x_i + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_i$$

$$y_{a,\,i} = a_i$$

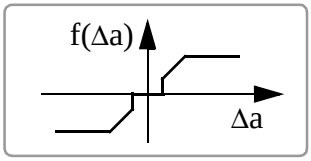
With the states, valve position a_i and valve velocity ω_i. The input signal u comes from the internal controller and $y_{a,i}$ is the measured position of valve i.

a = valve angle
$w = \dot{a}$ = valve angle ratio
u = voltage to run the servo DC-motor
y = measured position
i = valve index $\in \{15, 16, 18, 22\}$

The valve acts linear for small steps, less than 2-3 degrees, in the setpoint. For larger steps the internal controller signal saturates and the velocity reaches an upper limit. This is modeled with the internal controller law

$$u = f(\Delta a) = f(a_{i,R} - y_{a,i})$$

Here f(Δa) is a function to characterize saturation and dead zone of the internal controller.



$$u = f(\Delta a_i) = \begin{cases} 0 & ,|\Delta a_i| < 0{,}1 \\ k\Delta a_i & ,0{,}1 \le |\Delta a_i| \le 3 \\ u_{max}\,\mathrm{sgn}(\Delta a_i) & ,|\Delta a_i| > 3 \end{cases}$$

The physical influence on the air flow is modeled as an orifice with variable open area. The maximum open area $A_{i,m}$ is the only parameter that differs between different valve individuals. The open area A_i, is calculated from the valves position a_i, with the equation

$$A_i = A_{i,m}(1-\cos(a_i)) \tag{3.8}$$

The flow restriction has the same characteristics as the turbulent flow through an orifice. Equation 3.1 for calculating flow through an orifice is also used to calculate the flow through a valve.

$$M_i = \frac{K_0 A_i}{\sqrt{T}} \sqrt{P_u^2 - P_d^2}$$

The pressure P_u is the absolute pressure upstreams the valve and P_d the pressure downstream. The valve open area A_i is used as effective open area in the flow equation. The maximum effective open area A_m, can be found in the parameter list in Appendix A for each valve.

3.6.5 Ambient

In addition to the input signals discussed in previous chapters, ambient conditions also affect the performance of the Environmental Control System. Ambient conditions used in the principle model are pressure and temperature in the bleed air, P_{eng} and T_{eng} respec-

tively. Also the ambient pressure P_{amb} is used as input to the model. Some more or less necessary ambient conditions are not used in the model, among others, fighter velocity and air moisture can be mentioned. The velocity can not really be neglected so it is assumed to be constant in the simulations, representing a specific flight case.

3.7 Model errors

The larger a present fault is, the easier it is to detect. To detect small faults an accurate model is needed. With perfect knowledge of the 'real' system, arbitrary small faults can be detected. In order to increase reality to the problem formulation, errors are added to the model. Model parameters are distorted and noise is added to measured signals.

3.7.1 Parameter uncertainties

All parameters are randomly distorted. A script adds a random relative error to each parameter. The error is gaussian distributed and set to achieve a risk of 1% for a parameter deviation more than 10% of the nominal value. This gives a standard deviation of 0,0388.

If X is a stochastically distributed variable $X \sim N(0, \sigma)$ this means that X/σ is a standardized gaussian distribution with $X/\sigma \sim N(0, 1)$. To fulfill the requirements above we get a σ of

$$P(|X| > 0{,}1) = 2P(X < -0{,}1) = 2P\left(\frac{X}{\sigma} < \frac{-0{,}1}{\sigma}\right) = 1\%$$

the standard deviation σ, is calculated with the matlab expression

$$\sigma = \frac{-0{,}1}{\text{norminv}(0{,}005)} = 0{,}0388$$

All parameters, *p*, are distorted with this distribution according to the relation

$$p = p_0(1 + x)$$

where x is a sample from the mentioned distribution.

Model error is introduced when the parameters are distorted. The purpose with this is to make the diagnosis situation more realistic. In a real situation are model errors always present.

3.7.2 Sensor noise

Measuring in an highly electrical environment involves problems with sensor noise. To reflect a real situation sensor noise is added to all measured signals. In simulink, band-limited white noise blocks are used to distort the measured signals. The noise power is set relative the signal level to achieve a realistic signal behavior.

3.8 Fault Modeling

Faults mentioned in Chapter 3.2 are together with some additional faults modeled and added to the principle model. Each fault corresponds to a fault mode according to the list below.

3.8.1 Fault modes

Each modeled fault has a corresponding fault mode. Twelve possible faults are considered in the work, hence twelve fault modes are considered, these are:

$\boldsymbol{NF}$	No Fault
$\boldsymbol{FA_{leak}}$	Leakage in Avionics
$\boldsymbol{FV_{15j}}$	Valve 15 Jamming
$\boldsymbol{FV_{16j}}$	Valve 16 Jamming
$\boldsymbol{FP_{cp}}$	Cooling pack pressure sensor bias
$\boldsymbol{FM_{av}}$	Avionics flow sensor bias
$\boldsymbol{FM_{cab}}$	Cabin flow sensor bias
$\boldsymbol{FP_{av}}$	Avionics pressure sensor bias
$\boldsymbol{FP_{cab}}$	Cabin pressure sensor bias
$\boldsymbol{FV_{15p}}$	Valve 15 potentiometer feedback bias
$\boldsymbol{FV_{16p}}$	Valve 16 potentiometer feedback bias
$\boldsymbol{FP_{amb}}$	Ambient pressure sensor bias

Localization of the fault modes in the principle are visualized in Figure 17. The faults are two valve jamming errors, one leakage and 8 sensor faults.

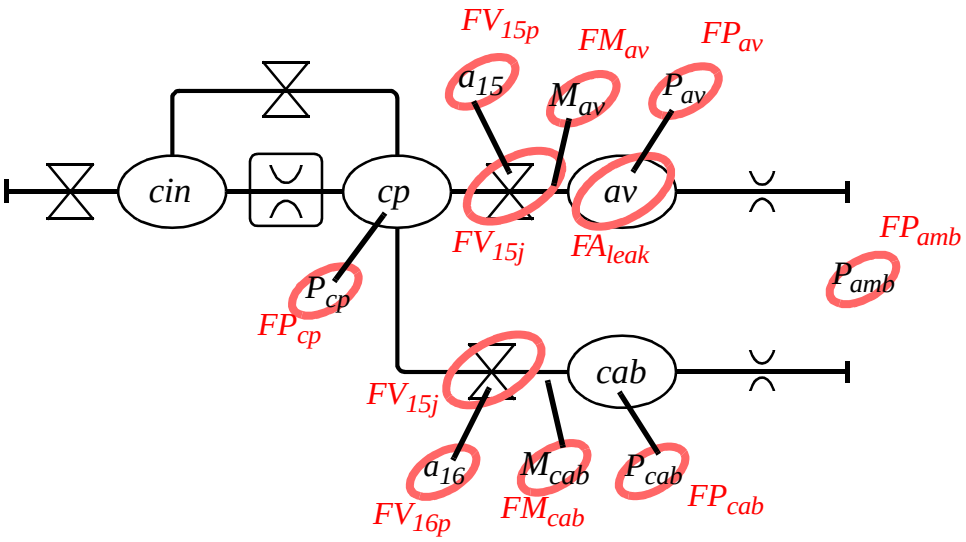


FIGURE 17. Localization of possible faults visualized in the principle model

3.8.2 Leakage

A leakage in the avionics compartment is modeled as an orifice with unknown area between the compartment and ambient air. This generates an additional flow out of the avionics compartment

$$M_{leak} = \frac{A_{leak} K_0}{\sqrt{T_{av}}} \sqrt{P_{av}^2 - P_{amb}^2}$$

where $A_{leak} = 0$ in the fault free case with no leakage and $A_{leak} \neq 0$ when a leakage occurs. The flow is added to Equation 3.4, which becomes

$$\dot{P}_{av} = \frac{RT_{av}}{V_{av}} (M_{av} - M_{ao} - M_{leak}) \tag{3.9}$$

3.8.3 Valve jamming

Valve jamming or valve stiction is a fault mode that characterizes the behavior of a valve that got stuck. It can be caused by increased static friction or deposits on the commuter of the servo motor. The model from valve position setpoint $a_{i,R}$, to valve position a_i is described in Chapter 3.6.4, here written as $a_i = g(t, a_{i,R})$.

$$a_i = \begin{cases} g(t, a_{i,R}(t)) & ,t < t_0 \\ g(t_0, a_{i,R}(t_0)) & ,t \geq t_0 \end{cases}$$

with $i \in \{15, 16, 18, 22\}$ and t_0 is the time when the fault occurs.

3.8.4 Valve potentiometer bias, F_{vp}

A valve potentiometer bias is modeled by a constant added to potentiometer feedback signal. The fault affects both servo controller feedback loop and measured signal fed to the diagnosis system. The feedback control will settle the valve in a position with zero setpoint error. The valve disc will however be in the wrong position. As a consequence, the fault can not be detected by using signals from position setpoint and measured position. The fault must be detected by the valves unexpected influence on the process.

$$y_{a15} = a_{15} + b_{a15}$$
$$y_{a16} = a_{16} + b_{a16}$$

with $b_i = 0$ in the fault free case with no leakage and $b_i \neq 0$ when a fault occurs.

3.8.5 Sensor bias

Bias is modeled in each sensor signal. This might seem as a strange fault at first glance, but it includes the special case with loose connections that very well might occur. Loose connections are often interpreted as a total loss of signal and not difficult to detect.

$$y_i = i + b_i$$

with $i \in \{P_{cp}, T_{cp}, M_{av}, T_{av}, P_{cab}, M_{cab}, T_{cab}, P_{av}\}$
with $b_i = 0$ in the fault free case and $b_i \neq 0$ when a fault occurs.

4 Diagnosis on the principle model

A diagnosis system is developed to exemplify different model based approaches to achieve fault detection and isolation. Two basic principles are used in the work, *estimation error principle* and *parameter estimation principle*. These are also modified to show the use of *adaptive thresholds* when model errors are known. Concepts as *decoupling* and *structured hypothesis tests* are discussed and exemplified.

During the development, a focus is held on ECS characteristics and known problems. The diagnosis system is built to detect a set of faults, which all originates from experience with the ECS. Data to feed the diagnosis system is produced by the principle model.

4.1 Test quantities

The diagnosis system is, in a first step, built up from 12 test quantities. After some performance evaluation an extra sensor is considered in the system and additional tests are built in step two, with a total of 17 test quantities.

Building of test quantities is restricted by available sensors and modeled relationships between measured quantities. Equations from the principle model are used to describe such signal relations in the fluid dynamics. The valve model is relaxed and a lower order model is used for valve relations in the diagnosis system.

The strategy used when building the test quantities, is to build as many as possible and afterwards select and keep the tests with best performance. A detailed discussion of how to automatize the selection of the best tests can be found in [11]. In this work the tests are chosen from the results of Monte Carlo simulations.

The first four test quantities, T_1 to T_4, uses the estimation error principle for cabin fluid relations. The tests T_5 to T_7 are based on the same fluid relations as the first four tests. By using a valve model they are slightly modified and different signals can be used to feed the test algorithm. Test T_7 and T_8 uses the estimation error principle with adaptive thresholds to diagnose valve dynamics. Tests T_{10} to T_{12} are used to test relations in avionics fluid relations. In step two of the diagnosis system, an extra sensor for measuring avionics pressure is used. This extra information is used to build five more tests, T_{13} to T_{17}, for avionics fluid relations.

In the following sections, Section 4.1.1 to Section 4.1.8, the test quantities are described in detail. Some notations are used to achieve a more compact and easily understood text. In several estimates an expression for the valves open area is needed, this is calculated from the valves position according to the relation

$$y_{A,\,i} = A_{i,\,0}(1 - \cos(y_{a,\,i}))$$

with the notation copied from Equation 3.8. Another convenient notation is the grouping of cabin and avionics fault modes in two different sets of fault modes.

$$R_{av} = \{FM_{av}, FA_{leak}, FV_{15j}, FV_{15p}\}$$
$$R_{cab} = \{FM_{cab}, FV_{16j}, FV_{16p}, FP_{cab}\}$$

This is suitable since cabin and avionics fault modes often are independent of each other.

4.1.1 T_1, estimation error

The first test quantity is built according to the estimation error principle applied to the air flow in to the cabin, M_{cab}. By using Equation 3.5 together with known signals, an estimate of the air flow can be calculated. Parameters in the equation are known and all quantities but temperature are measured. The temperature are approximated with a constant which will introduce some errors to the estimate. A temperature error of 30°C is within normal operating range and will distort the estimate less than 5%, which is tolerated.

From Equation 3.5 an expression with known signals applied, yields

$$\hat{\mathrm{M}}_{\mathrm{cab}} = \frac{\mathrm{y}_{\mathrm{A16}}\mathrm{K}}{\sqrt{\mathrm{T}}}\sqrt{\mathrm{y}_{\mathrm{Pcp}}^2 - \mathrm{y}_{\mathrm{Pcab}}^2}$$

from which the first test quantity is derived, as the estimation error

$$\boldsymbol{T_1} = \left| y_{Mcab} - \hat{M}_{cab}(y_{Pcp}, y_{Pcab}, y_{A16}) \right|$$

Test T_1 is used to test the first hypothesis H_1^0

$$H_1^0: \qquad F_p \in R_1 = \{NF, FP_{amb}, FV_{16j}, R_{av}\}$$

$$H_1^1: \qquad F_p \in R_1^c = \{FP_{cp}, FP_{cab}, FM_{cab}, FV_{16p}\}$$

4.1.2 T_2, estimation error with observer

Instead of directly measuring the cabin pressure in the equation above, an observer can be used to estimate it.

$$\dot{\hat{P}}_{cab} = \frac{RT_{cab}}{V_{cab}}\left(\hat{M}_{cab} - \frac{A_{cab}K}{\sqrt{T_{cab}}}\sqrt{\hat{P}_{cab}^2 - y_{Pamb}^2} + k_2(y_{Mcab} - \hat{M}_{cab}) \right)$$

this decouples the cabin pressure sensor signal and with a new combination of signals an estimate of the air flow can be achieved

$$\hat{M}_{cab2} = \frac{y_{A16}K}{\sqrt{T_{cp}}}\sqrt{y_{Pcp}^2 - \hat{P}_{cab3}^2}$$

leads to the second test quantity

$$\boldsymbol{T_2} = \left| y_{Mcab} - \hat{M}_{cab2}(y_{Pcp}, y_{Pamb}, y_{A16}) \right|$$

for testing the second hypothesis

$$H_2^0: \qquad F_p \in R_2 = \{NF, FP_{cab}, FV_{16}, R_{av}\}$$

$$H_2^1: \qquad F_p \in R_2^c = \{FP_{cp}, FP_{amb}, FM_{cab}, FV_{16p}\}$$

4.1.3 T_3 and T_4, observers driven by different sources

The two tests T_1 and T_2, above uses flow estimates to verify physical relationships. Given a physical model with the necessary information to get a relation between a set of signals, any of the signals can be extracted from the model and used as an estimate. Instead of estimating air flow, T_3 and T_4 are built to estimate cabin pressure. Each estimation driven by different observers for cabin pressure.

$$\boldsymbol{T_3} = \left| y_{Pcab} - \hat{P}_{cab}(y_{Pcp}, y_{Pamb}, y_{A16}) \right|$$

with estimate from observer

$$\dot{\hat{P}}_{cab} = \frac{RT_{cab}}{V_{cab}} \left(\frac{y_{A16}K}{\sqrt{T_{cp}}} \sqrt{y_{Pcp}^2 - \hat{P}_{cab}^2} - \frac{A_{cab}K}{\sqrt{T_{cab}}} \sqrt{\hat{P}_{cab}^2 - y_{Pamb}^2} + k(y_{Pcab} - \hat{P}_{cab}) \right)$$

used for testing the hypothesis

$$H_3^0: \qquad F_p \in R_3 = \{NF, FM_{16}, FV_{16p}, R_{av}\}$$

$$H_3^1: \qquad F_p \in R_3^c = \{FP_{cp}, FP_{amb}, FP_{cab}, FV_{16p}\}.$$

Test T_4, just like T_3, estimates the cabin pressure but with a different observer driven by another set of signals.

$$\boldsymbol{T_4} = \left| y_{Pcab} - \hat{P}_{cab}(y_{Pamb}, y_{M16}) \right|$$

from observer

$$\dot{\hat{P}}_{cab} = \frac{RT_{cab}}{V_{cab}} \left(M_{cab} - \frac{A_{oc}K}{\sqrt{T_{cab}}} \sqrt{\hat{P}_{cab}^2 - y_{Pamb}^2} + k(y_{Pcab} - \hat{P}_{cab}) \right)$$

for testing of the fourth hypothesis

$$H_4^0: \qquad F_p \in R_4 = \{NF, FP_{cp}, FA_{16p}, FA_{16j}, R_{av}\}$$

$$H_4^1: \qquad F_p \in R_4^c = \{FP_{amb}, FP_{cab}, FM_{16}\}$$

4.1.4 T_5, T_6 and T_7, model order reduction

Instead of directly using the measured position of the valve. The setpoint can be used together with the valve model to estimate the position and use the estimate instead of potentiometer signal. This would easily modify the tests above containing a measure of valve position to three new test.

With a reduction of model order, the test becomes less computational. Time constants for the valve dynamics are less than the general time constant of the air dynamics. If the valve dynamics are not considered an approximation of the valves real position as the setpoint can be used to create three new test quantities.

Tests T_5, T_6 and T_7 are almost the same as T_1, T_2 and T_3 respectively, the only modification is that the valve potentiometer signal, $y_{a,i}$ is exchanged for the valve setpoint, $a_{i,R}$. Each test now have a new hypothesis to test, instead of decoupling the valve setpoint the potentiometer signal is decoupled.

$$\boldsymbol{T_5} = \left| y_{Mcab} - \hat{M}_{cab}(y_{Pcp}, y_{Pcab}, A_{16R}) \right|$$

$$H_5^0: \quad F_p \in R_5 = \{NF, FP_{amb}, FV_{16p}, R_{av}\}$$

$$H_5^1: \quad F_p \in R_5^c = \{FP_{cp}, FP_{cab}, FM_{cab}, FV_{16j}\}$$

$$\boldsymbol{T_6} = \left| y_{Mcab} - \hat{M}_{cab}(y_{Pcp}, y_{Pamb}, A_{16R}) \right|$$

$$H_6^0: \quad F_p \in R_6 = \{NF, FP_{cab}, FV_{16p}, R_{av}\}$$

$$H_6^1: \quad F_p \in R_2^c = \{FP_{cp}, FP_{amb}, FM_{cab}, FV_{16j}\}$$

$$\boldsymbol{T_7} = \left| y_{Pcab} - \hat{P}_{cab}(y_{Pcp}, y_{Pamb}, A_{16R}) \right|$$

$$H_7^0: \quad F_p \in R_7 = \{NF, FM_{16}, FV_{16p}, R_{av}\}$$

$$H_7^1: \quad F_p \in R_3^c = \{FP_{cp}, FP_{amb}, FP_{cab}, FV_{16j}\}$$

4.1.5 T_8 and T_9, adaptive thresholds

The valve model can be used to estimate a position from the setpoint. A model of reduced order is used since the knowledge of the characteristics of the real valve is not well known. Two tests, one for each of valve 15 and 16, are built with the technique of adaptive thresholds described in Section 2.3.3. For details of notation and valve model approximations see Section 2.3.3.

$$a_i(t) = G(s)a_{i,R}(t)$$

Valve 16

$$\boldsymbol{T_8} = \left| y_{a16} - \hat{a}_{16}(a_{16R}) \right|$$

$$J_{th8}(t) = k_8(\left|\Delta G(s)\hat{a}_{16R}(t)\right| + c_8)$$

$$H_8^0: \quad F_p \in R_8$$

$$H_8^1: \qquad F_p \in R_8^C = \{FV_{16p}, FV_{16j}\}$$

and valve 15

$$\boldsymbol{T_9} = \left|y_{a15} - \hat{a}_{15}(a_{15R})\right|$$

$$J_{th9}(t) = k_9(\left|\Delta G(s)\hat{a}_{15R}(t)\right| + c_9)$$

$$H_9^0: \qquad F_p \in R_9$$

$$H_9^1: \qquad F_p \in R_9^C = \{FV_{15p}, FV_{15j}\}.$$

4.1.6 T_{10} and T_{11}, estimation error

By taking a closer look at Equation 3.4 it is seen that all signals available in the cabin must be used to get a relation between the signals. All signals are needed to establish a relation in the model of the cabin fluids. It is impossible to decouple any signal but those never used. By adding the valve model, one more relation can be derived and the measured valve position can also be decoupled.

The avionics air flow is estimated as

$$\hat{\mathrm{M}}_{av} = \frac{\mathrm{y_{A15}K}}{\sqrt{\mathrm{T}}}\sqrt{\mathrm{y_{Pcp}^2} - \hat{P}_{av}^2}$$

using an observer for avionics pressure

$$\dot{\hat{P}}_{av} = \frac{RT_{av}}{V_{av}}\left(\hat{M}_{cab} - \frac{A_{ao}K}{\sqrt{T_{av}}}\sqrt{\hat{P}_{av}^2 - y_{Pamb}^2} + k_2(y_{Mav} - \hat{M}_{av})\right)$$

This gives the test quantity

$$\boldsymbol{T_{10}} = \left|y_{Mav} - \hat{M}_{av}(y_{Pcp}, y_{Pamb}, y_{A15})\right|$$

used for decision in the hypothesis test

$$H_{10}^0: \qquad F_p \in R_{10} = \{NF, FP_{cp}, FP_{amb}, FV_{15r}\}$$

$$H_{10}^1: \qquad F_p \in R_{10}^c = \{FP_{cp}, FP_{amb}, FM_{av}, FV_{15p}\}$$

T_{11} is a test similar to T_{10}, the same observer is used and the avionics air flow is estimated once again. The difference lies in how knowledge of valve position is gathered. The idea is the same as used when building T_5 to T_7, a valve model is added to the estimate algorithm and the valve setpoint can be used instead of the measured valve position.

By assuming that valve dynamics are faster than avionics fluid dynamics, the valve model can be relaxed to

$$\hat{A}_{15} = A_{15m}(1 - \cos(a_{15R}))$$

and the estimation of avionics air flow becomes

$$\hat{\mathrm{M}}_{av} = \frac{\hat{A}_{15}(a_{15R})\mathrm{K}}{\sqrt{\mathrm{T}}}\sqrt{\mathrm{y}_{\mathrm{Pcp}}^2 - \hat{P}_{av}^2}$$

$\hat{P}_{av}$ is calculated from the observer above, also used for calculating T_{10}. Now test T_{11} becomes

$$\boldsymbol{T_{11}} = \left| y_{Mav} - \hat{M}_{av}(y_{Pcp}, y_{Pamb}, a_{15R}) \right|$$

and it is used for taking decision in the hypothesis test

$$H_{11}^0: \qquad F_p \in R_{11} = \{NF, FV_{16p}, R_{cab}\}$$

$$H_{11}^1: \qquad F_p \in R_{11}^c = \{FP_{cp}, FP_{amb}, FM_{av}, FV_{15j}\}$$

4.1.7 T_{12}, parameter estimation with RLS

This test is built with the parameter estimation principle. No extra relations are used, the test is based on the same signals and the same models as test T_2. Instead of using the signals for estimating the air flow, they are used for estimating the leakage area in the avionics compartment.

$$\boldsymbol{T_{12}} = \left| \hat{A}_{leak}(y_{Pcp}, y_{Pamb}, y_{A15}, y_{Mav}) \right|$$

Model parameters are generally considered to be constant over time and therefore it is a good idea to give the estimation slow dynamics in order to reduce the influence of sensor noise.

For filtering the signal a Recursive Least Squares RLS, algorithm is chosen. For further reading regarding RLS algorithms see [2]. A weighted least squares algorithm can with a well chosen weighting sequence be made recursive. A recursive algorithm is well suited for on line diagnosis since it consumes a minimum of memory for signal history.

The weighted least squares algorithm

$$\hat{\theta}_t = \arg\min_{\theta} \sum_{k=1}^{t} \beta(t, k)[y(k) - \varphi^T(k)\theta]$$

with the weighting sequence

$$\beta(t, k) = \lambda(t)\beta(t - t, k), \qquad 1 \le k \le t - 1$$

$$\beta(t, t) = 1$$

gives us the estimate

$$\hat{\theta}(t) = \hat{\theta}(t-1) + L(t)[y(t) - \varphi^T(t)\hat{\theta}(t-1)]$$

$$L(t) = \frac{P}{\lambda + \varphi^T P \varphi}$$

$$P(t) = \left[P(t-1) - \frac{P\varphi\varphi^T P}{\lambda + \varphi P \varphi}\right] / \lambda$$

With this algorithm the leakage area can be estimated from Equation 3.4

$$\dot{\hat{P}}_{av} = \frac{RT_{av}}{V_{av}}\left(\frac{A_{15}K_0}{\sqrt{T_{cp}}}\sqrt{P_{cp}^2 - P_{av}^2} - \frac{(A_{oa} + A_{leak})K_0}{\sqrt{T_{av}}}\sqrt{P_{av}^2 - P_{amb}^2}\right)$$

rewritten for extracting values to feed the RLS algorithm

$$\underbrace{\frac{\dot{P}_{av}V}{RT} - \frac{A_{15}K_0}{\sqrt{T_{cp}}}\sqrt{P_{cp}^2 - P_{av}^2} + \frac{A_{oa}K_0}{\sqrt{T_{av}}}\sqrt{P_{av}^2 - P_{amb}^2}}_{y} = \hat{A}_{leak}\ \underbrace{\frac{K_0}{\sqrt{T_{av}}}\sqrt{P_{av}^2 - P_{amb}^2}}_{\varphi}$$

Using a forgetting factor of $\lambda = 0{,}99$ gives a response time of 100 samples, i.e. 3 seconds. The parameter estimation becomes

$$\hat{A}_{leak}(t) = \hat{A}_{leak}(t-1) + L(t)[y(t) - \varphi^T(t)\hat{A}_{leak}(t-1)]$$

Since the same relations and signals as test T_{11} are used, the hypothesis will also become the same

$$H_{12}^0: \qquad F_p \in R_{12} = \{NF, FV_{16p}, R_{cab}\}$$

$$H_{12}^1: \qquad F_p \in R_{12}^c = \{FP_{cp}, FP_{amb}, FM_{av}, FV_{15j}\}$$

4.1.8 T_{13} to T_{17}, extra sensor P_{av}

T_{13} to T_{17} are built with the use of an extra sensor for avionics pressure, P_{av}. This is an example of a strategic placed sensor that increases the analytical redundancy. By using this sensor the performance of the leakage area estimation is increased and four extra tests can be built.

With the use of a sensor for avionics pressure instead of an observer, a more accurate estimate of the leakage area is obtained.

$$\underbrace{\frac{\dot{P}_{av}V}{RT} - \frac{A_{15}K_0}{\sqrt{T_{cp}}}\sqrt{P_{cp}^2 - P_{av}^2} + \frac{A_{oa}K_0}{\sqrt{T_{av}}}\sqrt{P_{av}^2 - P_{amb}^2}}_{y} = \hat{A}_{leak}\ \underbrace{\frac{K_0}{\sqrt{T_{av}}}\sqrt{P_{av}^2 - P_{amb}^2}}_{\varphi}$$

$$\boldsymbol{T_{13}} = A_{leak}(t) = \hat{A}_{leak}(t-1) + L(t)[y(t) - \varphi^T(t)\hat{A}_{leak}(t-1)]$$

The avionics air flow can be estimated without the use of an observer

$$\hat{\mathrm{M}}_{av} = \frac{\mathrm{y_{A15}K}}{\sqrt{\mathrm{T}}}\sqrt{\mathrm{y_{Pcp}^2 - y_{Pav}^2}}$$

$$\boldsymbol{T_{14}} = \left| y_{Mav} - \hat{M}_{av}(y_{Pcp}, y_{Pav}, y_{A15}) \right|$$

When introducing the extra sensor, enough redundancy is obtained to use the estimation error principle on avionics pressure

$$\dot{\hat{P}}_{av} = \frac{RT}{V_{av}}\left(\frac{y_{A15}K}{\sqrt{T}}\sqrt{y_{Pcp}^2 - \hat{P}_{av}^2} - \frac{A_{ao}K}{\sqrt{T}}\sqrt{\hat{P}_{av}^2 - y_{Pamb}^2} + k(y_{Pav} - \hat{P}_{av})\right)$$

$$\boldsymbol{T_{15}} = \left| y_{Pav} - \hat{P}_{av}(y_{Pcp}, y_{Pamb}, y_{A15}) \right|$$

Yet another estimate of avionics pressure can be obtained

$$\dot{\hat{P}}_{av} = \frac{RT}{V_{av}}\left(M_{av} - \frac{A_{ao}K}{\sqrt{T}}\sqrt{\hat{P}_{av}^2 - y_{Pamb}^2} + k(y_{Pav} - \hat{P}_{av})\right)$$

$$\boldsymbol{T_{16}} = \left| y_{Pav} - \hat{P}_{av}(y_{Pamb}, y_{Mav}) \right|$$

The last test quantity T_{17}, uses both avionics and cabin fluid relations. The cooling pack outlet pressure P_{cp}, is estimated from two different sets of signals and compared to each other.

$$\hat{P}_{cp1} = \sqrt{T\left(\frac{M_{av}}{A_{15}K}\right)^2 - y_{Pav}^2}$$

$$\hat{P}_{cp2} = \sqrt{T\left(\frac{M_{cab}}{A_{16}K}\right)^2 - y_{Pcab}^2}$$

$$\boldsymbol{T_{17}} = \left| \hat{P}_{cp1}(y_{Pcab}, y_{A16}, y_{Mcab}) - \hat{P}_{cp2}(y_{Pav}, y_{A15}, y_{Mav}) \right|$$

4.2 Thresholds

Thresholds can be chosen in many different ways. In this work, the thresholds are given values calculated from Monte Carlo simulations. The Monte Carlo simulations are evaluated with two different approaches, discussed in Section 2.3. The first approach, called *maximum deflection*, tolerates no false alarms and becomes very restrictive. This implies a bad fault detection ability. In order to increase the diagnosis performance all tests are inspected with the second approach, *histogram inspection*. In some cases a lowered

threshold is found to increase the diagnosis performance. In these cases the thresholds are lowered.

4.2.1 Maximum deflection

As mentioned earlier, the *maximum deflection* method is a straightforward way to obtain values of thresholds. From a large set of simulations, the threshold are set to not fire in any case. In this evaluation of the diagnosis system performance, 55 simulations are used to automatically set the thresholds. It would be desirable to use more simulations but it is a time consuming method since each simulation takes about 1 hour. Instead 20% is added on top of the peak test quantity value to obtain a margin for false alarms. An example with thresholds set after 5 simulations can be seen in Figure 18.

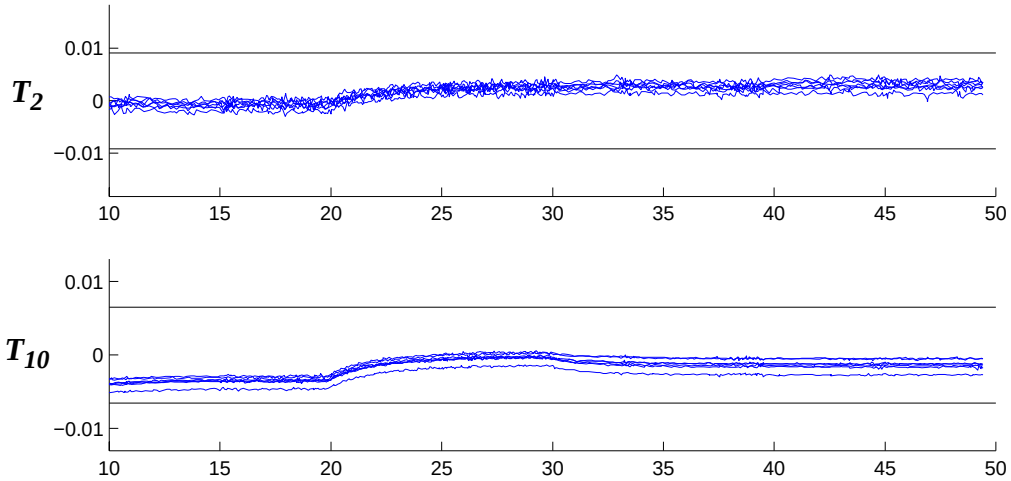


FIGURE 18. Thresholding with the maximum deflection approach

This approach to find thresholds is straightforward and systematic, easy to automatize. But there is a risk of being too restrictive. If the thresholds are set too high, they will never fire, not even when a fault occurs.

4.2.2 Histogram inspection

A more realistic approach is to set the limit as a compromise between risk of false alarm (threshold to low) and risk of not detecting a fault (high threshold). This is the idea with the *histogram inspection* approach.

A correct built test quantity is zero for all decoupled fault modes and high for all others. By looking at a histogram of the test quantity deviation after a set of simulations it might occur that it is worth lowering the threshold in order to increase the possibility to detect faults. In Figure 19 an example can be seen. The figure shows the evaluation of test quantities from 12 simulations with different fault modes. The upper histogram shows a test

quantity when the present fault mode is decoupled. The lower histogram is the same test quantity when the present fault mode is not decoupled.

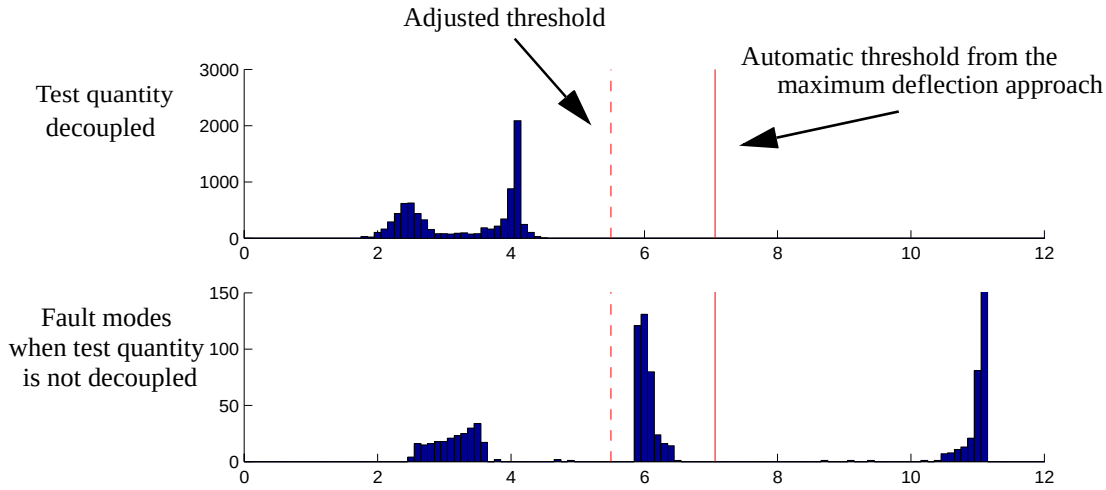


FIGURE 19. Histogram inspection

4.2.3 Tuning of adaptive thresholds

The adaptive thresholds in Section 4.1.5 for test quantity T_8 and T_9 can not be chosen with the same procedure as the rest of the thresholds. All other thresholds have just one parameter to tune, the threshold level itself. The adaptive threshold

$$J_{th}(t) = k(|\Delta G(s)a(t)| + c)$$

has two parameters to tune, k and c. The procedure used for tuning the adaptive thresholds involves more work. First the parameter k is set to a value of 1. In a static condition $G(s) = 0$ and with $k = 1$, c becomes an upper limit of the static prediction error. It is possible to choose the value c with the same procedure as mentioned above, as long as no change in the valves setpoint occurs. When a value for c is obtained, the value k becomes the threshold gain during transients. From simulations, with a fault free system, with steps of different magnitude in valve setpoint, k is chosen to not fire alarms. The simulated steps are supposed to reflect normal valve operation.

4.3 Decision logic

A summary and overview of the hypothesis tests is put together in the decision structure in Table 1. Here it is shown how each test relate to the fault modes.

Table 1: Decision structure

	NF	FP_{cp}	FP_{amb}	FP_{cab}	FM_{16}	FA_{16}	FA_{16j}	FM_{15}	FA_{15}	FA_{15j}	FA_{leak}	FP_{av}	
T_1	0	X	0	X	X	X	0	0	0	0	0	0	Step 1
T_2	0	X	X	0	X	X	0	0	0	0	0	0	
T_3	0	X	X	X	0	X	0	0	0	0	0	0	
T_4	0	0	X	X	X	0	0	0	0	0	0	0	
T_5	0	X	0	X	X	0	X	0	0	0	0	0	
T_6	0	X	X	0	X	0	X	0	0	0	0	0	
T_7	0	X	X	X	0	0	X	0	0	0	0	0	
T_8	0	0	0	0	0	X	X	0	0	0	0	0	
T_9	0	0	0	0	0	0	0	0	X	X	0	0	
T_{10}	0	X	X	0	0	0	0	X	X	0	X	0	
T_{11}	0	X	X	0	0	0	0	X	0	X	X	0	
T_{12}	0	X	X	0	0	0	0	X	X	0	X	0	
T_{13}	0	X	X	0	0	0	0	X	X	0	X	X	Step 2
T_{14}	0	X	0	0	0	0	0	X	X	0	0	X	
T_{15}	0	X	X	0	0	0	0	0	X	0	X	X	
T_{16}	0	0	X	0	0	0	0	X	0	0	X	X	
T_{17}	0	0	0	X	X	X	0	X	X	0	0	X	

A decoupled fault mode, marked with a 0, can not affect the test. The fault modes marked with an X might affect the test.

This structure is divided in to two steps. The first step uses a diagnosis system fed with sensor signals available in the ECS today. Step 2 is a diagnosis system with an additional sensor in the avionics compartment.

For each test the decoupled fault modes in R_k contributes with a 0 in the table and each fault mode in the complement R_k^C contributes with an X. The purpose with the hypothesis testing is to obtain a diagnosis statement. The diagnosis statement is a list of possible fault modes for the observed process. Each hypothesis test contributes with a set of possible fault modes S_k, with

$$S_k = \begin{cases} \Omega, \text{ if test } k \text{ is quite} \\ R_k^C, \text{ if test } k \text{ fires an alarm} \end{cases}$$

These are combined to a diagnosis statement S, by taking the intersection of all sets S_k.

$$S = \bigcap_k S_k$$

The final statement S is a list of possible fault modes that can explain the process behavior.

4.4 Evaluation of diagnosis system

Evaluation of the diagnosis system is performed with simulations of the principle model running in different fault modes. System inputs and measured signals from the principle model is fed to the diagnosis system. The signals are used in the diagnosis system to calculate the test quantities which all are calculated simultaneously. The typical behavior of the test quantities are shown in Figure 20. The figure shows a simulation with the fault mode FP_{cab} present. The diagnosis system is turned off the first seconds to give the observers some time to stabilize. After three seconds the diagnosis system is turned on and after five seconds the fault occurs. On the right side of the figure it is marked which of the test quantities that might fire and which are decoupled, all according to the incidence structure in Table 1.

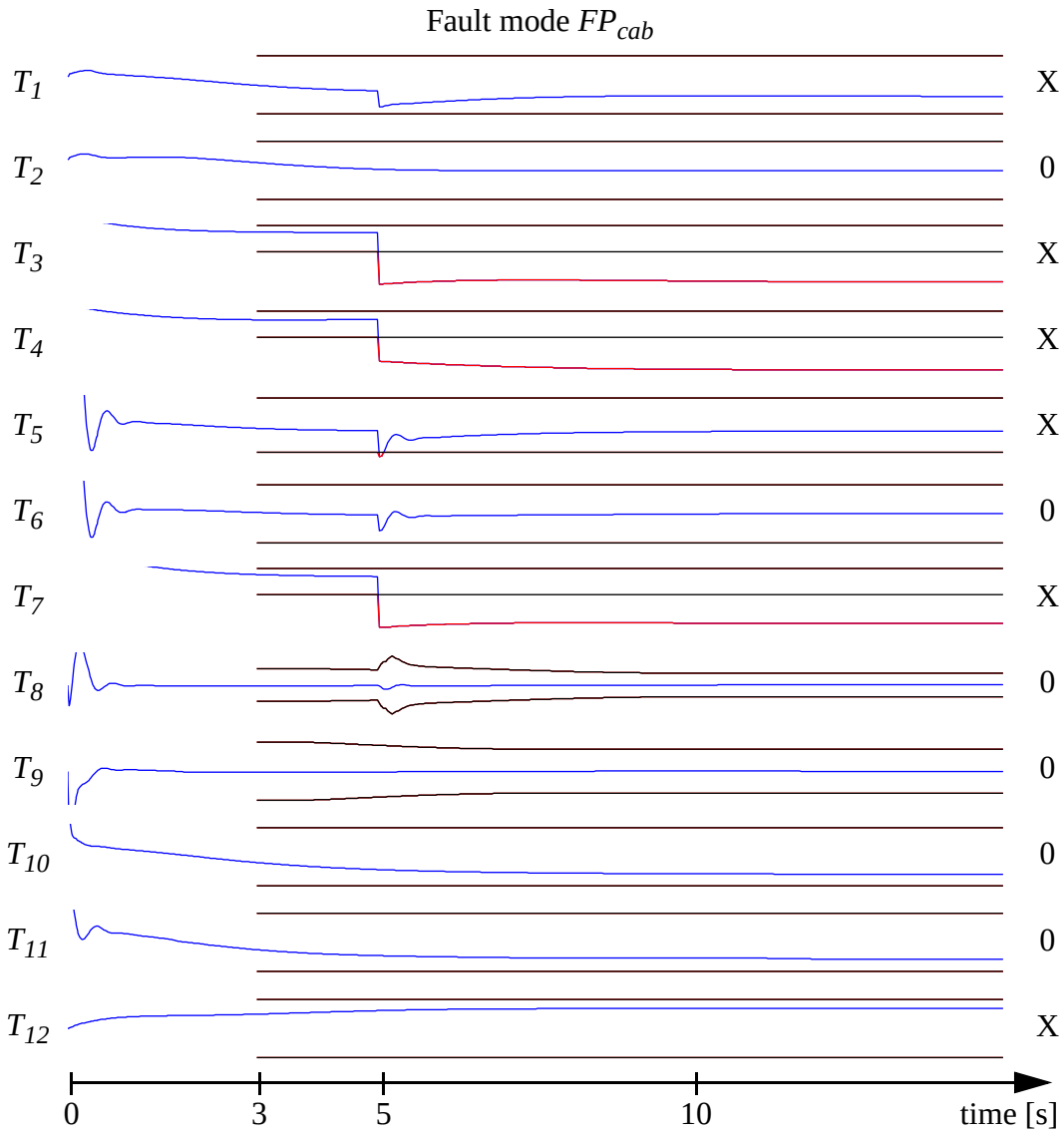


FIGURE 20. Typical test quantity behavior

The diagnosis statement is instant in time and some considerations about when to generate an alarm must be made. It is not always suitable to fire an alarm immediately when a fault statement occurs. A small probability for taking the wrong decision is present. In order to lower the risk of false alarm further, a delay of the alarm is built so it does not fire until consecutive samples of fault statement occurs.

Two diagnosis statements are generated as a summary of each simulation. The first statement consider only one sample of the simulation, and fire immediately when an alarm is detected. The second uses a time window of 10 seconds and accumulates all firing test quantities during that time, the diagnosis statement becomes the result from all firing tests during the last 10 seconds.

Two different diagnosis systems are evaluated. The first uses only sensor signals available in the ECS today. This means that only tests T_1 to T_{12} in Section 4.1 are used. In Table 1 these are found in the upper part of the table, step 1.

The second diagnosis system is an extension of the first. It uses an additional sensor for the pressure in the avionics compartment. Here all tests in Section 4.1 are used to take the diagnosis statement. All tests in Table 1 are used, both those in step 1 and step 2.

For evaluation of the diagnosis systems they are applied to simulations with the principle system. Simulations are performed with each of the possible fault modes present, one at the time.

In Table 2 the first diagnosis system is seen. A simulation for each fault mode is made and sensor signals available today are used to feed the diagnosis system.

Table 2: Diagnosis statement with available sensors

	in a time window (all simulation)	at an instant (last sample)
NF	no larm	no larm
FP_{cp}	$\{FP_{cp}\}$	$\{FP_{cp}\}$
FP_{amb}	$\{FP_{amb}\}$	$\{FP_{amb}\}$
FP_{cab}	$\{FP_{cab}\}$	$\{FP_{amb}, FP_{cab}\}$
FM_{cab}	$\{FM_{cab}\}$	$\{FM_{cab}\}$
FV_{16p}	$\{FV_{16p}\}$	$\{FV_{16p}\}$
FV_{16j}	$\{FV_{16j}\}$	$\{FV_{16j}\}$
FM_{av}	$\{FP_{cp},FP_{amb},FM_{av},FA_{leak},\}$	$\{FP_{cp},FP_{amb},FM_{av},FA_{leak}\}$
FV_{15p}	$\{FV_{15p}\}$	$\{FV_{15p}\}$
FV_{15j}	$\{FV_{15j}\}$	no larm
FA_{leak}	$\{FP_{cp},FP_{amb},FM_{av},FA_{leak}\}$	$\{FP_{cp},FP_{amb},FM_{av},FA_{leak}\}$

Most fault modes are perfectly isolated. Only two fault modes can not be isolated, these are leakage in avionics compartment FA_{leak}, and avionics mass flow sensor bias FM_{av}. It is also worth to notice the difference between the two diagnosis statements. The first statement based on a time window, perfectly isolates the fault in the ambient pressure sensor FP_{amb}, this is not the case with the second statement based on an instant sample. The fault mode with valve 15 jamming FV_{15j}, is only detected with the first statement, the reason to

this is seen in Figure 21, showing how the jamming valve is a fault that is only detected intermittent.

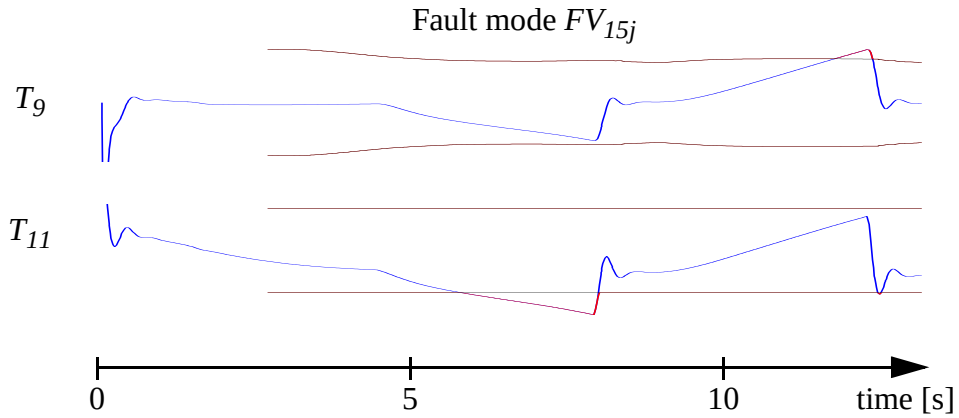


FIGURE 21. Intermittent fault detection

In Table 3 the evaluation of the second diagnosis system is seen. The second diagnosis system is the same as the first but extended with five additional tests made possible by the additional pressure sensor.

Table 3: Diagnosis statement with extra sensor P_{av}

	in a time window (all simulation)	at an instant (last sample)
NF	no larm	no larm
FP_{cp}	$\{FP_{cp}\}$	$\{FP_{cp}\}$
FP_{amb}	$\{FP_{amb}\}$	$\{FP_{amb}\}$
FP_{cab}	$\{FP_{cab}\}$	$\{FP_{amb}, FP_{cab}\}$
FM_{cab}	$\{FM_{cab}\}$	$\{FM_{cab}\}$
FV_{16p}	$\{FV_{16p}\}$	$\{FV_{16p}\}$
FV_{16j}	$\{FV_{16j}\}$	$\{FV_{16j}\}$
FM_{av}	$\{FM_{av}\}$	$\{FM_{av}\}$
FV_{15p}	$\{FV_{15p}\}$	$\{FV_{15p}\}$
FV_{15j}	$\{FV_{15j}\}$	no larm
FA_{leak}	$\{FP_{amb}, FA_{leak}\}$	$\{FP_{amb}, FA_{leak}\}$
FP_{av}	$\{FP_{av}\}$	$\{FV_{15p},FP_{av}\}$

The performance of the second system is found to be increased. All fault modes are isolated except leakage in the avionics compartment. The leakage, FA_{leak} can not be isolated from ambient pressure sensor bias, FP_{amb}.

5 Discussion and conclusions

A model based diagnosis system has been designed and applied to a principle model of the Environmental Control System. The design procedure can roughly be divided into four major steps:

1. Gathering knowledge. The approach with model based diagnosis relies on knowledge of the supervised process. In order to obtain this knowledge a lot of documentation about the ECS have been gathered and studied. Not only knowledge about the supervised system is needed but also knowledge about possible faults and how they affect the system behavior. Information about possible and common faults are gathered by interviewing people with experience from the ECS.

2. Building the principle model. When designing a diagnosis system, reliable models of both process and possible faults are needed. In this work, a new model had to be developed to obtain desired functionality as dynamic behavior and possibility to simulate different fault modes.

Models in general are built with the intention to perform a certain kind of simulation. The existing model of the ECS, for example, is built to simulate static conditions of the system. In this work a dynamic model was desired, since one major advantage with model based diagnosis is the ability to handle transients by using dynamic models. Also the affect of faults, important to supervise, had to be included in the model.

A principle model was built for simulation of main behavior and functionality of the ECS. Parameter uncertainties and sensor noise were added to the model to make the simulations more realistic. Faults, known to occur in the ECS, were modeled, together with one additional hypothetical fault for the exemplification of diagnosis methods.

3. Designing the diagnosis system. The design of a diagnosis system consists mainly of building a number of tests. Each test uses a test quantity and a threshold for taking a decision. The decision assigns the present fault mode to a subset of all fault modes considered in the work.

The diagnosis system developed for the principle model is focused on supervision of the distribution part of the system. The same signals as those available in the real ECS, are taken from the principle model and fed to the diagnosis system. The diagnosis system is built by designing test quantities, using available signals and mathematical relations from the principle model. The strategy used when building the test quantities is to build as many as possible, and after evaluation select and keep those who increases the diagnosis performance.

4. Evaluation of diagnosis performance. The performance of the diagnosis system is evaluated with simulations of the system in different working conditions and with different faults present. Common requirements are discussed. Requirements considered in this work are low false alarm rate, low missed detection probability and a wish to isolate each fault mode. One common requirement not considered in this work is the mean time for detection.

Beside evaluation of the diagnosis system, a survey of existing diagnosis methods have been performed. Model based diagnosis have potential to add desirable functionality and increase performance on the existing diagnosis system. Advantages with a model based diagnosis system is discussed and compared to the existing system.

5.1 Model based diagnosis in the real ECS

A model based approach has potential to increase diagnosis performance and add functionality to the system. One great advantage with this approach is the structure for isolation of faults. Isolation of faults, especially online isolation, opens a number of possibilities for adding desirable functionality such as isolation of intermittent faults and improved redundancy management.

Another advantage, compared to the traditional approaches used today, is the possibility to handle supervision during system transients. Today, supervision during system transients is not working satisfactory. Some supervision are even shut off during transients due to a high false alarm rate. By using dynamic models to explain the process behavior, it is possible to handle system transients in the diagnosis system.

A model based diagnosis system, as the one used in this thesis, can with advantage, be used together with an existing diagnosis system. The general structure with a set of hypothesis tests makes it open for integration with other diagnosis approaches.

5.1.1 Isolation offline

The procedure to find the cause of an alarm, isolation, are today performed manually by a skilled technician. With system knowledge, information of logged data and records of alarms, it is possible to manually localize the fault. With automatic fault isolation this work would be made easier and more cost effective.

It is not reasonable to believe that an automatic system would perfectly isolate all possible faults, such a system would be too expensive to develop. The intention with an automatic system must be to reduce the work load for manual fault localization. Even a less advanced diagnosis system isolates some faults, and when it is not possible for the diagnosis system to perfectly isolate a fault, the diagnosis statement becomes a list of possible faults. In both cases, manual fault localization are made easier.

5.1.2 Online isolation

Situation awareness is an important concept in a fourth generation aircraft. With perfect knowledge of aircraft condition it is possible to use it for maximum possible performance. An online diagnosis system with isolation capabilities will contribute to the situation awareness. The Gripen fighter has capability to run in different modes depending on available functionality. This is implemented according to two concepts, graceful degradation and redundancy management.

When a system is malfunctioning it is often possible to shut it down completely or only parts of it. The concept of *graceful degradation* has the objective to minimize loss of functionality in a malfunctioning system. For example, if a valve in the ECS is jammed it might be possible to shut down only parts of the ECS and not all the system. With graceful degradation the aim is to shut down only parts of a faulty system that are malfunctioning.

Many critical systems has some kind of backup mode or auxiliary system for preventing severe damage in the case of failure. *Redundancy management* is the possibility to govern such modes and systems. Decisions to shut on or off redundant and main systems are taken by the pilot or an automatic system. In both cases it is important to know the actual condition of the systems for taking a correct decision. A well designed diagnosis system with online isolation can provide this information.

One last advantage with online diagnosis is the immediate isolation of a fault when it occurs. Some faults only occurs intermittent or with special conditions only present during flight. It might be difficult or impossible to provoke the same behavior on the ground at a service occasion. This makes some faults really difficult to isolate manually.

5.1.3 System saturation

An alarm is not only a warning to signal that something is wrong, it should also come with a recommendation to the pilot of suitable measures to take. This gives a somewhat peculiar situation as the diagnosis is implemented today.

The diagnosis system mainly supervises system output, an alarm is generated when the desired output is not achieved. An approach with limited possibilities to isolate the source of an alarm. If the desired output is not achieved it does not necessarily mean that the system is malfunctioning, in some situations it can be the result of system saturation.

The system is designed to work within some working conditions, if these are not fulfilled it can not produce the desired output. For example, the ECS can only deliver a limited amount of cooling air, especially at high altitudes with the engine running at a low thrust. An alarm caused by low flow of cooling air at such working condition should not generate the same alarm as an alarm from a system malfunction.

When the system is saturated, it is only a question of manage it correctly and it is up to the pilot to run the aircraft in such a way that the system is within the predefined working conditions. An alarm caused by system malfunction has a completely different message to the pilot, run carefully and let a qualified technician have a look at the system.

With a model based diagnosis system that isolates the source of an alarm, it is possible to give the pilot a correct recommendation, not only an alarm.

5.1.4 System transients

A well known problem today is the lack of possibility to supervise system transients. A model based diagnosis system is well suited for this task. The models used for designing the diagnosis system also decides the performance and functionality of the diagnosis system. If it is possible to build a reliable dynamic model of the ECS it is also possible to handle transients in the system. The model does not necessarily have to be very accurate but must catch the most important dynamics in the system.

5.2 Developing the principle model and the diagnosis system

To exemplify model based diagnosis methods, a process to supervise is needed. The principle model was built to produce data with the same characteristics as the ECS. This data is used to feed the diagnosis system built in this thesis. The principle model is has the capability to simulate a number of faults of special interest to supervise. Some of the faults

are studied since they are more common and some because they affect the safety of the aircraft,

5.2.1 The principle model

The purpose with the principle model is to catch the main dynamics of the ECS. Of course it would be interesting to verify the model, but it is a work that takes a lot more time than fits within the frames of this thesis. Model verification is a work big enough for a thesis on its own.

Some parts are most likely to not diverge from the reality. Especially the distribution to avionics has potential to be accurate. This part of the system consists only of static components as pipes and volumes, therefore it should be easier to model. The model of the cabin distribution is a little more rough since the outlet is modeled as an orifice but in reality it consists of a pressure controlling valve.

The cooling turbine probably introduce the most errors to model. It is modeled as a heat exchanger only, with no dynamics in the fluid flowing through it. In reality the flow through the turbine is affected by turbine speed and the spinning mass will introduce a time delay that probably is of the same size as other time constants in the model. If the model has to be more accurate a first step can be to model the dynamics of the flow through the turbine.

In the principle model not all control loops are modeled. The last control loop, for cabin temperature is omitted to speed up the simulations. The affect of this model reduction is probably not very large, since the change in cabin temperature is relatively small. A more precise diagnosis system maybe should add this loop to make the model more accurate.

5.2.2 Fault modeling

Building models of faults is a work quite characteristic for the design of model based diagnosis systems. Knowledge of different faults and their affect on the process is necessary for designing the diagnosis system. Most of the faults modeled in this thesis are faults that actually has occurred in the ECS.

Two faults are modeled in the valves, potentiometer bias and valve jamming. The fault with potentiometer bias is realistic and easy to model. The measured signal of a valves position is superimposed with a constant signal. This fault happens if the disc of the valve is not mounted correctly or if the potentiometer slides out of position.

The jamming of the valve occurs when the valve is worn out and the static friction increases, the fault is modeled as increased static friction occurring on a given time. The static friction is not present when the valve is moving, therefore the valve is stopped at the time when the fault occurs. This is made just to be sure that the fault actually is present in the system, to give the diagnosis system a fair chance to detect the fault.

Sensor bias is a fault modeled in all sensor signals, maybe not the most likely fault for all sensors. Different sensors are constructed in different ways and get different kinds of faults. A temperature sensor is not likely to get the same fault signature as a pressure sensor. The question of how to model faults for different sensors is open for future investigations.

There exist one fault that is the same to all sensors. All sensors are connected to a computer by a wire, if this wire is not connected properly the measured signal will be faulty. Many systems are designed to give a loose connector a known maximum or minimum value. Such a fault is trivial but important to supervise. It is not considered in this work due to the simplicity of such supervision, still it is important to supervise in a real system since it is known to happen.

The fault mode with leakage in avionics compartment is not important to supervise in a the real ECS. Leakage has occurred only once during last years and is added to the model to show the principle with parameter estimation.

5.2.3 Thresholds

Setting of the thresholds is the part of the work that decides the final performance of the diagnosis system. In this thesis, the settings are performed by evaluation of monte carlo simulations. To achieve a reliable, well tuned, diagnosis system it is necessary with more simulations than performed here. In a real application it is even realistic to assume that the thresholds must be tuned also after implementation of the diagnosis system.

5.2.4 Performance

The principle model, incorporating models of faults is supervised by a diagnosis system. The diagnosis system is mostly based on the same models as the ones used for simulation of the ECS, only the valves are modeled different. Outputs from the model and parameters in the principle model are distorted to give a realistic situation with model uncertainties.

In the diagnosis system, models of fluid dynamics are the same as in the principle model. Temperatures are at some points approximated with constants since they are known to not change much. Valves in the diagnosis system are modeled with a lower order model than in the principle model.

The diagnosis system is built with sensors available in the ECS today. With the setup used, it is possible to isolate all fault modes but leakage and bias in flow sensor to avionics. These faults can not be isolated from three other fault modes. With an extra sensor added for pressure in avionics, only the leakage can not be perfectly isolated. When using the extra sensor, leakage can not be isolated from only one other fault mode.

5.3 Continuation of the work

A continuation of the work has some delicate tasks to deal with. A collection of possible things to do is listed below.

5.3.1 Principle model verification

The principle model is built with intention to behave like the real ECS. It is built with parameters taken from the existing model in Easy 5. It would be interesting to see how well the model corresponds to the real system and see if it can be adjusted to use for a rough simulation of the ECS.

With a modified principle model of the ECS that corresponds to the real system, it would be possible to run the proposed diagnosis system on real data. The principle model is an simplification of the ECS and can not get very accurate, but also rough models can be used

for diagnosis, as long as they are reliable. For example the valve model used in the diagnosis system in this thesis is a rough model of the supervised valve and still it is possible to detect present faults. In the same way a reliable rough model of the ECS has potential to be used with success for building a diagnosis system of the real ECS.

5.3.2 Principle model improvements

If effort is put in validating the model some parts of the principle model must be improved. The most important is the energy transfer in the heat exchanger. How much heat that is taken from the air in the cooling pack is highly dependent of altitude, velocity and surrounding temperature. In the principle model all these parameters are considered at a fixed working condition with a constant heat transfer. A model of this relation is needed to run the principle model at different working conditions.

The flow through the cooling turbine is modeled as any flow between two volumes. The mass of the turbine will introduce dynamics affecting the flow through it. It is likely that this will cause a time constant of the same order as the ones modeled in the principle model. If this is the case, a model of that dynamic would increase the reliability of the principle model.

A last improvement that maybe is not that important is the model of the outlet from the cabin. In the principle model this is modeled as an orifice with constant area, in the real ECS this outlet is a valve controlling cabin pressure. This valve has a small operating range but it is not constant. The model error introduced here is somewhat compensated by parameter uncertainties applied to the principle model in the simulations. Each simulation will be performed with a different open area and after many simulation all valve positions has been simulated. Distribution of the parameter for the valves area is defined to cover the valves operating range. The task to control cabin pressure is performed by valve at cabin inlet.

5.3.3 Fault model improvements

The faults modeled in the principle model are of interest to supervise in the real ECS. By experience, most of the faults are known to actually occur and are desirable to supervise. The fault mode with leakage in avionics compartment is not needed in a real implementation of a model based diagnosis system. A leakage has occurred only once during the last five years and was detected without causing any hazardous situation.

The valve jamming is the fault that has been of most interest to supervise. In a near future all valves are exchanged to a new which possibly will not be worn out. This points out another dilemma, is it worth spending resources on a diagnosis system or should all effort be put in building a reliable system that never malfunctions? A compromise will probably give the best result.

Another fault related to valves is the possibility to get increased friction that would slow down the valve. This has not been considered to be a fault in the valves used today, but it is from a general point of view when supervising valves. When the valves are replaced with new ones, this might be a fault mode of interest.

In an attempt to build a diagnosis system for the real ECS, the fault mode with sensor bias should be replaced with a fault mode with sensor cutoff or shortcut. This is easier to detect

and makes the diagnosis system more reliable. The fault mode with leakage can probably be ignored to further increase the reliability of the diagnosis system.

5.3.4 Model completions

The bypass pipe for control of cabin temperature is not modeled in the principle model. The affect of this model order reduction is uncertain, it is possible that this part of the system also should be modeled. This would add the last of the five control loops to the model. Diagnosis performance does not improve just by using a more complete model, so the use of such model completion is uncertain.

5.3.5 Diagnosis system improvements

The diagnosis system built in the thesis, is focused on air dynamics at the distribution part of the ECS. It can be completed with test quantities detecting faults upstreams in the model, not only at the distribution parts. It can also be completed with diagnosis using temperature relations, not considered in this diagnosis system.

Appendix A: Model parameters

Parameters in the principle model are, when possible, taken straight from the existing Easy5 model[8]. Since the two models differ as much as they do, a straight translation is not always possible. Some parameters had to be made up, partly by using Easy5 as base and partly by simulations to get reasonable outputs. Parameters in the Easy5 model is a mix of US- and ISO- units, in the principle model all are converted to SI.

Easy5The principle model

Cabin:

Outlet diameter = 2inch	A_{co} = 26cm^2, using 15cm^{2*}
Volume, diameter = 40.5cm	
Volume, length = 9.84m	V_{cab} = 5.1m^3

Avionics:

Volume diameter = 2inch,	
Volume length = 5ft	V_{av} = 3.1dm^3
Outlet effective area= 1.1inch^2	A_{ao} = 7.1cm^2

Heat exchanger:

Actually the heat exchanger is used to model the cold air unit that consist of several parts. Turbine, secondary heat exchanger, condenser, water separator, reheater and pipes. These together make a volume of 13.5ft^3 and it is the volume used in the model. Or actually that volume is divided into two parts, inlet and outlet of heat exchanger. Between those two volumes a flow restriction and the heat exchanger is put. Maximum heat transfer is not known but a value is taken to fit the model.

Many small volumes that sum up to 13.5ft^3.	Total volume = 0.2dm^3
Inlet volume	V_{cin} = 0.11dm^3
Outlet volume	V_{cp} = 0.11dm^3
Maximum heat transfer	q_m = 15 kW

Valves**:

14HA, (27)	diameter = 1.46inch	A_{14m} = 10.8cm^2
15HA, (21)	diameter = 2.46inch	A_{15m} = 30.7cm^2
16HA, (73)	diameter = 2.46inch	A_{16m} = 30.7cm^2
18HA, (18)	diameter = 1.46inch	A_{17m} = 10.8cm^2
22HA, (59)	diameter = 2.46inch	A_{18m} = 30.7cm^2

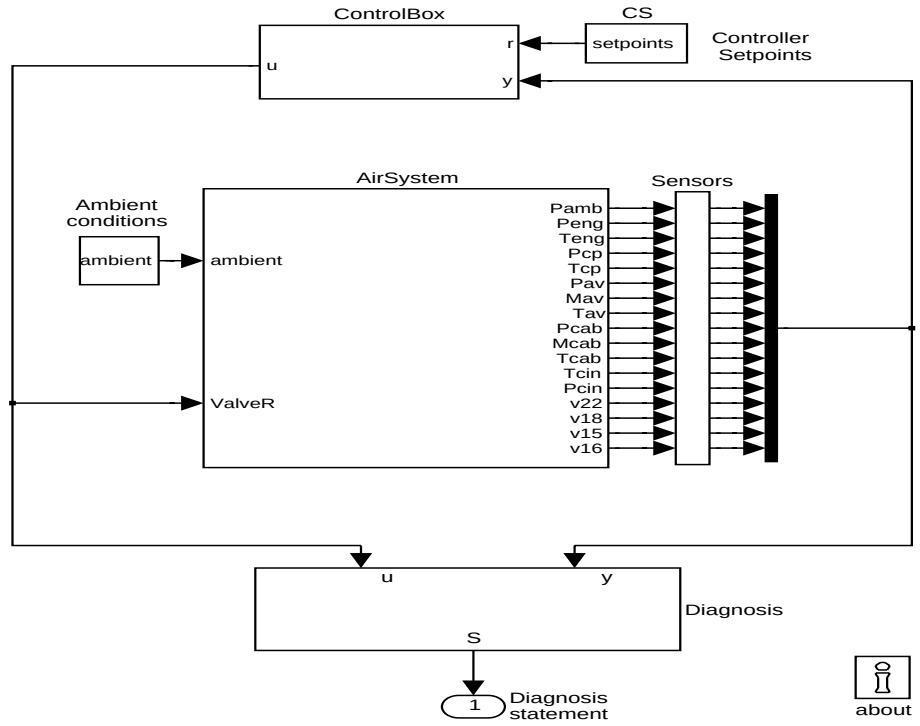
*. Outlet is actually a pressure controlled valve, set to keep pressure in cabin. In the principle model a constant orifice area is used, with half the area of the fully opened real valve.

**.Saab notation, (Hymatic notation)

Appendix B: The principle model in Simulink

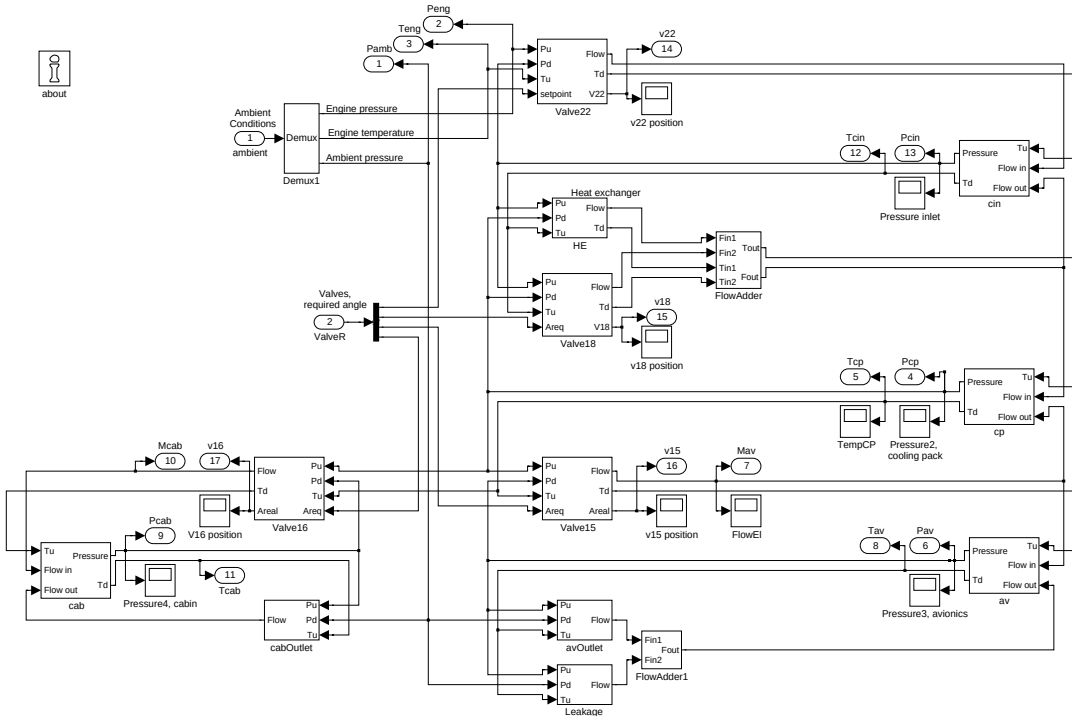
A schematic view of the simulation setup. In the middle of the figure the environmental process is seen with input and output signals. At the top of the figure the controlbox for running the four control loops and at the bottom the diagnosis system connected to available signals.

Simulation setup, overview	file:~ECSmodel/ecs3/ecs3.mdl

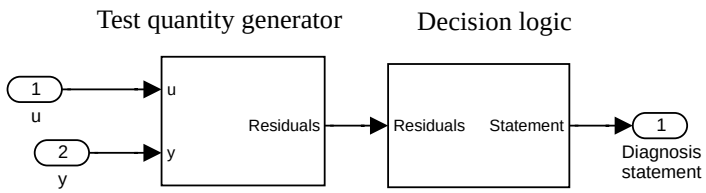


In order to shorten the evaluation time of the test quantities the model was divided in two parts. The first simulates the environmental process and the second runs the diagnosis system on the output from the first part.

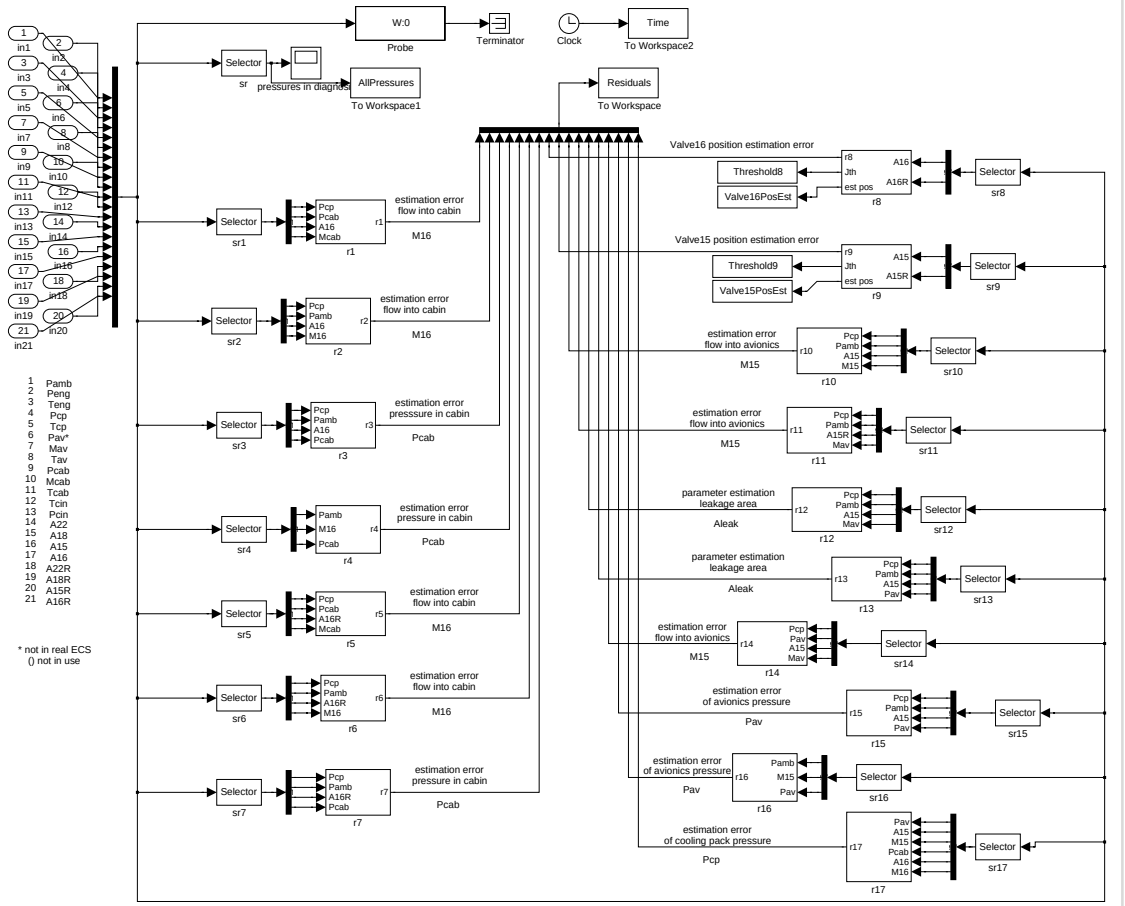
Environmental Control System | model: ecs3/AirSystem



Diagnosis System | model: ecs3/Diagnosis



Structured hypothesis tests | file:~ECSmodel/ecs3/Diagnosis2/Diagnosis2.mdl



Bibliography

[1]. Heat Transfer, J.P. Holman, McGraw-Hill1989
[2]. System Identification - theory for the user, Lennart Ljung, Prentice-Hall 1987
[3]. Simulering och Modellbygge, Lennart Ljung and Torkel Glad, Studentlitteratur 1991
[4]. Diagnosis and Supervision of Technical Processes, Mattias Nyberg and Erik Frisk, ISY Lith
[5]. FPL39, Utbildningsdokumentation, grundflygplan, Luftförsörjning, Saab AB, M5800-390011
[6]. Underhållshandbok F, UHF 36 Fpl39, J1-A.36-00-00-00A-412A, FMV
[7]. Hymatic, ECS documentation
[8]. Easy5 model, file: JAS39B_ECS_AC_3.mf.0
[9]. Kvalitet, Bo Bergman and Bengt Klefsjö, Studentlitteratur 1995
[10].Sannolikhetsteori och statistiskteori med tillämpningar, Gunnar Blom, Studentlitteratur 1989
[11].Model Based Fault Diagnosis, Mattias Nyberg, 1999
[12].On Integrity Monitoring of Integrated Navigation Systems, Jan Palmqvist, 1997