Bayesian networks for identifying incorrect probabilistic intuitions in a climate trend uncertainty quantification context

Hanea, A.M.; Nane, G.F.; Wielicki, B.A.; Cooke, R.M.

DOI
10.1080/13669877.2018.1437059

Publication date
2018

Document Version
Accepted author manuscript

Published in
Journal of Risk Research

Citation (APA)
Hanea, A. M., Nane, G. F., Wielicki, B. A., & Cooke, R. M. (2018). Bayesian networks for identifying incorrect probabilistic intuitions in a climate trend uncertainty quantification context. *Journal of Risk Research*, 1-16. https://doi.org/10.1080/13669877.2018.1437059

Bayesian Networks for identifying incorrect probabilistic intuitions in a climate trend uncertainty quantification context

A.M. Hanea[a] * G.F. Nane[b] B.A. Wielicki[c] R.M. Cooke[d]

[a] *Centre of Excellence for Biosecurity Risk Analysis, University of Melbourne, Australia*

[b] *Department of Applied Mathematics, Delft University of Technology, The Netherlands*

[c] *NASA Langley Research Center, Hampton, VA, USA*

[d] *Resources for the Future, Washington, DC, USA*

*Correspondence to: A.M.Hanea, CEBRA, University of Melbourne, Parkville, VIC 3010, Australia, email: anca.hanea@unimelb.edu.au

Bayesian Networks for identifying incorrect probabilistic intuitions in a climate trend uncertainty quantification context

Abstract

Probabilistic thinking can often be unintuitive. This is the case even for simple problems, let alone the more complex ones arising in climate modelling, where disparate information sources need to be combined. The physical models, the natural variability of systems, the measurement errors and their dependence upon the observational period length should be modelled together in order to understand the intricacies of the underlying processes. We use Bayesian networks (BNs) to connect all the above mentioned pieces in a climate trend uncertainty quantification framework. Inference in such models allows us to observe some seemingly nonsensical outcomes. We argue that they must be pondered rather than discarded until we understand how they arise. We would like to stress that the main focus of this paper is the use of BNs in complex probabilistic settings rather than the application itself.

1 Introduction

Probabilistic thinking is not always as intuitive as one would like, but it is the only proper way to avoid pitfalls and misconceptions, and to extract useful information from conflicting signals in the data. Many interesting phenomena may occur when *classical* statistical intuition plays tricks on the thought process. Some of these are: negative learning, concordant disagreement, discordant agreement, shifting consensus and obsolescence (Cooke and Wielicki 2016a).

Negative learning contradicts the familiar paradigm that new observations always increase our confidence, and welcomes the idea that an unexpected piece of information can sometimes increase the uncertainty. In a multivariate normal setting the conditional variance is always less than or equal to the unconditional variance, but this is rather atypical. Under a more complex error model unexpectedly high results may influence the prior enough to increase variance.

Concordant disagreement happens when disagreement decreases uncertainty, rather than increasing it, contradicting the intuition that when conflicting results occur, one must be *right*, and the other must be *wrong*. Rather than discarding one of the measurements, they can be combined in a probabilistic model acknowledging that conditional on conflicting results, we expect the errors to be negatively correlated.

Discordant agreement contradicts the intuition that different measurements of same system, returning the same values, should yield the same conclusions. This is not necessarily the case when the variance affects the mean, causing difference between more and less accurate measurements, even when they return the same value.

It is hard to imagine what combining concordant measurements will result in, apart from strengthening confidence in the common result. *Shifting consensus* happens when by combining the measurements their joint variance decreases, and that allows the posterior mean to drop.

A natural question that emerges in various engineering settings is if it is worthwhile retaining older systems once better systems have become available. Older measuring systems may be considered *obsolete* once new, improved ones become available. However it sometimes turns out that the older systems can still provide useful information, when used in conjunction with the new systems.

This paper will only touch on three of these phenomena, i.e. negative learning, discordant agreement and obsolescence, and it will do so using a climate change application. The focus of this paper in not the application itself, but the probabilistic modelling tools which allow us to tackle the above mentioned misconceptions. These tools are Bayesian networks (BNs). BNs are probabilistic graphical models which allow for rigorous multivariate uncertainty modelling, while providing a relatively simple visualization of the complicated relationships among the variables. This last feature makes them great tools for communicating future risks and their implications. We use a particular type of BNs, namely Non-Parametric Bayesian Networks (NPBNs) to help us understand where some of the disagreements on this topic are coming from, and how to use these disagreements to advance science.

The remainder of the paper is organized as follows: Section 2 describes the climate trend application, its modelled and measured variables, and it reveals how various sources of uncertainty in climate trends materialize in the dependence relationships among these variables. Section 3 provides background on NPBNs in general, and it discusses the NPBNs built for this application. The analysis addressing emerging challenges in understanding the complex probabilistic system is contained in Section 4. We draw conclusions in Section 5.

2 Climate trend modelling

The adverse effects of climate change are recognised by scientists all over the world. Environmental and social changes are constantly being observed and future changes depend on climate policies and emission paths. The decision making context we are interested in here is one where a switch is required from a current to a reduced emissions path. We switch to a reduced emission path when a trigger variable exceeds a threshold value with requisite confidence. The benefits of such a switch depend on the natural variability of the modelled processes and the uncertainty of measuring instruments. An enhanced measuring system might reduce uncertainty and lead to better predictions. However, given the different sources of uncertainty and the complexity of the problem, a value of information analysis is required to quantify the true benefits of the new observing system relative to the existing one. Such analysis is undertaken in Cooke et al. (2014) and Cooke et al. (2016). For the analysis undertaken here we consider three variables, two of which are measurable: the decadal rate of global temperature rise (DTR) and the decadal percentage change in cloud radiative forcing (CRF). DTR and CRF can be measured with different measurement systems which will differ in the technology used, hence they will have different measurement errors.

An enhanced measuring system that uses better calibration of existing systems to observe trends is the Climate Absolute Radiance and Refractivity Observatory (CLARREO) system[1]. CLARREO is set to launch no earlier than 2020.

[1]https://clarreo.larc.nasa.gov/

This may be contrasted with existing systems: for global temperature rise, these are weather satellite infrared spectrometers **I**ASI (EUMETSAT instrument), **A**IRS (NASA instrument), and **C**rIS (NOAA instrument), abbreviated as IAC. They look at about 1/3 of the Earth's emitted infrared radiation where CO_2 and H_2O absorb radiation in varying levels. For CRF, the existing system is the Clouds and the Earth Radiant Energy Systems (CERES) system, a broadband radiation budget instrument. It measures the total reflected solar energy as a single value, and the total emitted thermal infrared energy as a second value.

Along with the measured values for DTR and CRF, their theoretical values may be calculated using deterministic models. These theoretical values are determined by Equilibrium Climate Sensitivity (ECS). ECS is defined as the equilibrium change in global surface air temperature as a response to doubled CO_2. One such determinstc model is the Dynamic Integrated model of Climate and the Economy (DICE) assessment model (Nordhaus and Sztorc 2013), certified by the Inter-Agency Working Group on the Social Cost of Carbon IWG SCC (2013).

Putting all of the above together (variables of interest, their modelled versions and their measured versions) results in a model involving seven variables: ECS, the theoretical DTR, the theoretical CRF, two measured DTR (IAC and CLARREO), and two measured CRF (CERES and CLARREO). Despite the DICE model being deterministic these variables are nonetheless uncertain. Different sources of uncertainty have been identified to account for the climate trend uncertainty; a list of key examples, as well as relevant references can be found in Cooke et al. (2014) and Cooke et al. (2016). An important source of uncertainty is introduced by the natural variability, denoted by σ^2_{var}, which is mainly driven by oceanic and atmospheric dynamics of the climate systems, along with solar variability. Degrading calibration of satellite instruments over time yields additional uncertainties, denoted by σ^2_{cal}. Finally, we consider σ^2_{orbit}, the satellite orbit sampling uncertainty. Following Leroy et al. (2008) and Wielicki et al. (2013), the total uncertainty in a decadal trend δm can be decomposed as:

$$(\delta m)^2 = \frac{12}{(\Delta t)^3}\left(\sigma^2_{var}\tau_{var} + \sigma^2_{cal}\tau_{cal} + \sigma^2_{orbit}\tau_{orbit}\right), \tag{1}$$

where Δt is the observation period length (in years) and τ_{var} is the autocorrelation characteristic time length of natural variability, accounting for the dependencies between successive measurements. Similarly, τ_{cal} and τ_{orbit} denote the characteristic time lengths of calibration and satellite orbit uncertainty. Note that δ_m decreases at a rate proportional to $(\Delta t)^{-3/2}$. The decomposition from equation (1) can be used in general for any climate variable of interest, and the units will be determined by the choice of the variable. A derivation of equation (1) can be found in the Supplementary Online Material of Cooke et al. (2014). It suffices here to say that the variance decomposition for linear models is the basis of the uncertainty representation in equation (1).

Using the σ^2_{var} and σ^2_{cal} from equation 1, we can describe the uncertainty of the variables by adding noise to their deterministic values obtained from the DICE model. The relationship between these variables can be depicted in a graphical form which reflects the flow of influence between them. The chosen graphical representation is shown in Figure 1.

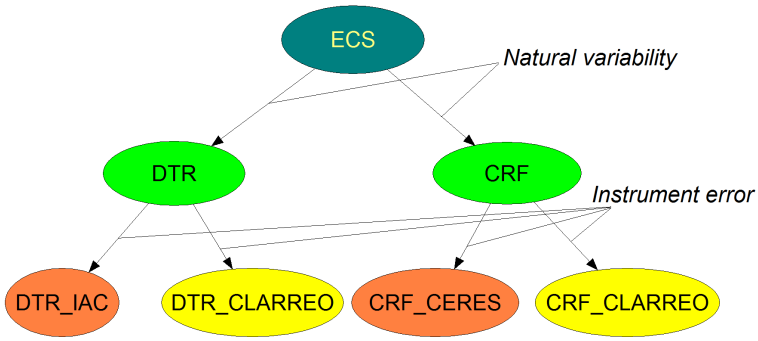


Figure 1: Graphical representation of the joint distribution of the variables of interest.

The functional relationships (modelled with DICE) between ECS and DTR, and between ECS and CRF are augmented with uncertainty by adding noise which describes the natural variability (see the upper half of Figure 1). Instrument error affects the DTR and CRF variables, and these errors are modelled as additive noise to obtain the DTR and CRF as measured by the current observational system, and by the enhanced system (see the lower half of Figure 1).

3 Bayesian Networks for climate modelling

The graphical model presented above can be thought of as a Bayesian network (BN)(Pearl 1988). A BN consists of a directed acyclic graph (DAG) and a set of conditional distributions. The DAG alone is sometimes called the qualitative part of the BN and the set of conditional distributions is considered to be the quantitative part of such models. Each node in the graph represents a random variable and the arcs represent direct potential dependence[2] relationships. The absence of arcs guarantees a set of conditional independencies. The direct predecessors of a node are called parents, and the direct successor are children. A marginal distribution is specified for each node with no parents, and a conditional distribution is associated with each child node. The (conditional) distributions serve as the quantitative information about the strength of the dependencies between the variables involved. The graph with the conditional independence statements encoded by it, together with the (conditional) distributions, represent the joint distribution over the random variables represented as nodes in the graph. BNs whose nodes represent discrete random variables are called discrete BNs. These models specify marginal distributions for the nodes with no parents, and conditional probability tables for child nodes. However, many domains require reasoning about the behaviour of a mixture of discrete and continuous variables. These domains are often called hybrid domains. Hence, BNs involving both discrete and continuous variables are called hybrid BNs. Purely continuous BNs are often assumed to represent a fully parametric joint distribution (e.g. multivariate normal). This assumption is often too restrictive and it can be relaxed by allowing for the factorisation of the joint distribution into marginals and dependence.

In this research we use BNs to tackle the difficulties of integrating disparate information sources in an ECS estimation problem. Because all the variables involved in our analysis are measured on a continuous scale, we chose to work with a Non - Parametric Bayesian Network (NPBN). NPBNs (Hanea et al. 2015) associate nodes with random variables for which no parametric marginal distribution assumption is made and arcs with one-parameter conditional copulae (**?**), parameterized by rank

[2]The strength of such relationship becomes apparent only through quantification.

correlations.

3.1 Non - Parametric Bayesian Networks

NPBNs concepts were proposed in Kurowicka and Cooke (2004) and extended and refined in Hanea et al. (2006) and Hanea et al. (2010). NPBNs construct the joint distribution of a set of variables represented as a DAG by coupling marginal distributions of all variables with the dependence structure constructed from bivariate pieces of dependence.

NPBNs associate nodes with random variables for which no parametric marginal distribution assumption is necessary, and arcs with one-parameter conditional copulae (**?**), parameterised by Spearman's rank correlations. The (conditional) copulae are assigned to the arcs of the NPBN according to a protocol that depends on a (non-unique) ordering of the parent nodes. The conditional copulae, together with the one-dimensional marginal distributions and the conditional independence statements implied by the graph uniquely determine the joint distribution, and every such specification is consistent (Hanea et al. 2015).

The marginal distributions can be obtained from data or experts (Cooke 1991). Even though the empirical marginal distributions are used in most cases, parametric forms can be also fitted. The (conditional) copulae used in this method are parametrised by constant (conditional) rank correlations that can be calculated from data or elicited from experts (Morales et al. 2008). The conditional rank correlations need not (in general) be constant. Nevertheless they are usually assumed constant for convenience. Each parent-child influence is associated with a (conditional) rank correlation. For each child with more than one parent, the set of respective parents can be ordered in a non-unique way. When experts participate in building the structure of the NPBN they can agree upon an ordering of the parents that corresponds to their decreasing strength of influence on the child. Otherwise, the order of the parents can be driven by the availability of data. The rank correlation assigned to the arc between the child and its first parent (in the chosen ordering) is an unconditional one. The rack correlation between the child and its second parent is conditional on the first parent in the ordering; the rank

correlation between the child and the third parent is conditional on the first two parents, and so on.

The rank correlation has a number of attractive properties, among which: it measures monotone dependence, rather than just linear dependence, and it is independent of the marginal distributions.

We employ the NPBN software called Uninet. Uninet was designed by the Department of Applied Mathematics of the Delft University of Technology and licensed by LightTwist Software. A free version is available for academic users at http://www.lighttwist.net/wp/. Initially developed for the Dutch Ministry of Transport, Uninet was designed for mixed (discrete and continuous) multivariate distributions in very high dimensions (Ale et al. 2009). Uninet uses the normal copula to realise the rank correlations which quantify the arcs. This allows for extremely fast exact inference. The reasons for this along with some properties of the normal copulae and the partial and conditional correlations are outlined in the Appendix.

3.2 Quantification of the NPBN

The starting point for the BN modelling and analysis in this paper is the graph from Figure 1, which represents the DAG of a NPBN. It consists of seven nodes representing continuous random variables and six arcs. To quantify this model we need seven continuous distributions, and six rank correlations. No variable has more that one parent, and as a consequence, there is no conditional rank correlation needed for this model.

The variable of interest is ECS. Its probability distribution has been derived in Roe and Baker (2007), truncated for an increased consistency with IWG SCC 2013 (Cooke and Wielicki 2016b), and presented below:

$$ECS = \frac{1.2}{1-f}, \qquad \text{where } f \sim N(0.62, 0.19^2). \tag{2}$$

The ECS distribution exhibits a right long tail that suggests higher probabilities for high and very high temperature changes rather than for small or no temperature change. It is easy to sample from the ECS distribution (using equation 2), and then

to obtain the corresponding *true* CRF and the *true* DTR values, using the DICE model. Then natural variability noise is added to each of the possible true trend values (using σ^2_{var} from equation 1). These form the empirical distributions of the variables DTR and CRF when the natural variability is taken into account.

Instrument error is added to the CRF and DTR values already perturbed by natural variability, to obtain the empirical distributions of the measured quantities. The σ 's used for the measurement errors are the ones corresponding to the noise contribution from calibration and orbit sampling. The values of these parameters that have been used in this and previous analyses are taken from Table 2 of Cooke et al. (2014), for CLARREO and IAC global temperature trend uncertainty, and from Table 5 of Cooke and Wielicki (2016b), for the variance decomposition for CLARREO and CERES. They are reproduced below, in Table 1, for completeness.

Table 1: Natural variability and observation uncertainties for CRF and DTR

	CRF		DTR	
	CLARREO	CERES	CLARREO	IAC
σ_{var}	0.6	0.6	0.08	0.08
τ_{var}	0.8	0.8	1.5	2.3
σ_{cal}	0.15	1	0.03	0.18
τ_{cal}	10	10	5	5
σ_{orbit}	0.21	0.006	0.019	0.012
τ_{orbit}	1	1	1	1

The dependence structure of the seven dimensional distribution is determined by the rank correlations corresponding to the arcs of the NPBN, and the choice of the copula, which in this case, is the normal copula. The next step in quantifying the model is to assign rank correlations to the arcs of the NPBN. The normal copula inherits many of the normal distribution properties, including the relationship between the rank and the product moment correlation given by Pearson's transformation (see Appendix). The time dependent standard deviations may be used to compute product moment correlations in the following manner: assume we observe the random variable X with random noise ξ, with X and ξ independent (which is how we modelled the natural variability and the instrument errors). Then the product moment correlation of X and $Z = X + \xi$ is given by:

$$\rho_{X,Z} = \frac{\sigma_X}{\sqrt{\sigma_X^2 + \sigma_\xi^2}}. \tag{3}$$

These product moment correlations can be then transformed to the rank correlations needed to quantify the arcs of the NPBN. A different approach is an empirical one: since we have the empirical bivariate margins of ECS and DTR, ECS and CRF, and all modelled and measured versions of DTR and CRF, the empirical rank correlations are calculated from the corresponding sampled bivariate distributions. It is worth mentioning that instead of calculating the rank correlation between the CRF disturbed by natural variability and ECS, we calculate the rank correlation between the CRF disturbed by natural variability and the true CRF. Theoretically, these two rank correlations are equal since the true CRF and ECS have a monotonic relationship and rank correlations are invariant under monotone transformations of the margins.

There is no correlation between the uncertainties of the CERES and IAC instruments. There is almost no common technology, and their calibration issues are very different[3].

Assuming CLARREO launch in 2020, the system can be observed for a large number of years after that. We can build several models which reflect the situation in 2030, 2050, etc. The qualitative part of these models (the graphs) will stay the same, but the quantification will change, and these changes will be reflected in the dependence coefficients rather than in the marginal distributions. Inferring decadal trends over a shorter period of time (e.g. 10 as opposed to 30 years) results in higher contributions from natural variability and measurement noise (see equation 1). This will be reflected in lower correlations. In time, less variability and less measurement noise are translated into larger correlations between the variables

[3]The enhanced measuring system for CRF employs a reflected solar spectrometer with a large 2D detector array (512 by 512 detectors) that uses scans of the sun, moon, and nearby deep space to do calibration and SI traceability. It shares no types of components with the IR spectrometer, uses a 2-axis gimbal to point the entire instrument so that the exact same optics path is used for solar, lunar, and earth viewing observations. The enhanced measuring system for temperature is an interferometer that uses deep cavity blackbodies (0.9998 emissivity where 1.0 is perfect), 3 different temperature phase change cells to calibrate temperature of the blackbody to SI standards, a blackbody emissivity monitor, and varies its blackbody temperatures for calibration from 200K to 320K. The physics of how such instruments would change in orbit has no common element, even the electronics of these instruments are very different.

in the NPBN models.

Figure 2 shows two NPBNs. The NPBN from the left (right) hand side reflects the situation after ten (30) years of observing decadal trends that follow the launch of the enhanced measuring system in 2020. In both models the nodes are now depicted as the empirical marginal distributions of the variables, together with the expected values and the standard deviations (presented at the bottom of each of the nodes). The rank correlations associated with the arcs of the model on the left hand side are substantially lower than the ones associated with the arcs of the model on the right hand side.

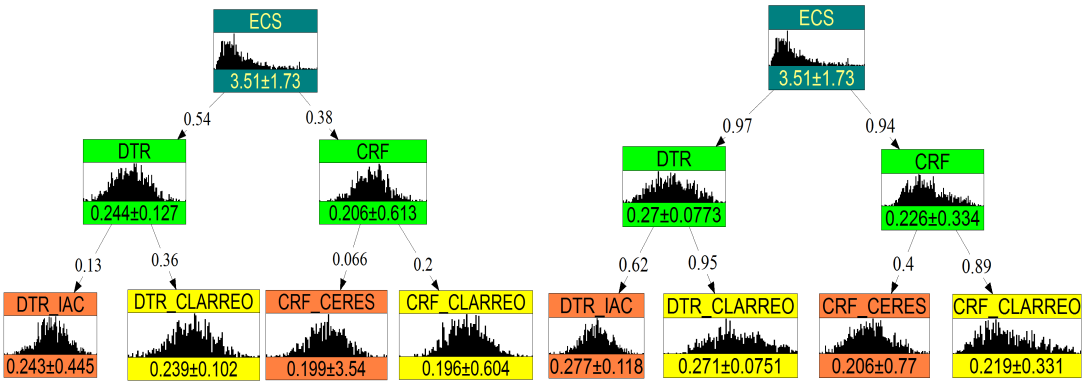


Figure 2: NPBN after 10 years (left) and after 30 years (right) of observing decadal trends that follow the CLARREO launch in 2020.

Instead of considering separate slices of time as different NPBNs as above, we can include the time dimension into one temporal NPBN which is usually called a Dynamic NPBN (DNPBN). A DNPBN has extra arcs which correspond to the dependence between time frames. This allows variables to be connected with (dependent on) themselves at earlier stages in time. For the DNPBN in this paper we consider only the variable DTR with its two corresponding measuring systems (IAC and CLARREO), for which we include three time steps from 2030, 2050, and 2070. The choice of excluding CRF is made for clarity of presentation, rather than a methodological restriction. Adding explicit dependence (arcs) between the different temporal versions of DTR induces dependence between the corresponding measuring systems.

The new (temporal) dependencies represented as new arcs in the DNPBN need to be quantified using conditional rank correlations, since each DTR node except

DTR30 has now two parents instead of one (as in the static version of the models). For example the node DTR50 has ECS and DTR30 as parents. The arc between ECS and DTR50 is quantified by the (unconditional) rank correlation used in the static model, whereas the arc between DTR30 and DTR50 is quantified by the conditional rank correlations of the two variables given the first parent (i.e. ECS). This conditional rank correlation is calculated using the unconditional correlations of each pair of the three variables mentioned above in addition to the formula connecting rank and the product moment correlation, the relationship between the partial and conditional product moment correlations (under the normal copula assumption),and finally a recursive formula to calculate partial correlations (see the Appendix for the exact formulae). The unconditional rank correlation between DTR30 and DTR50 is calculated empirically through simulations of the bivariate marginal distribution of DTR30 and DTR50.

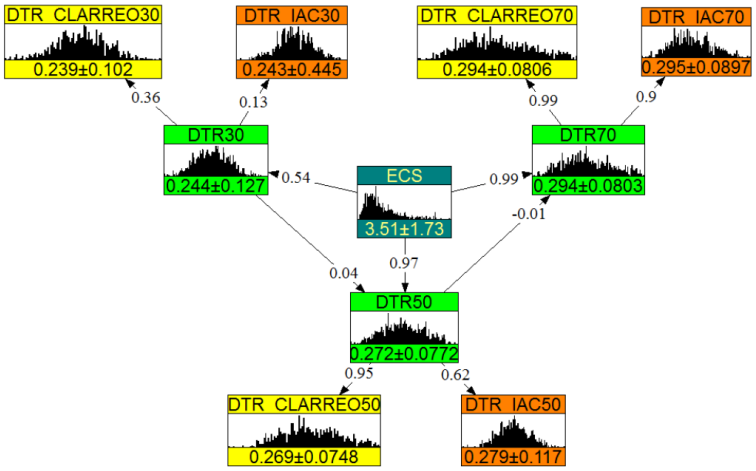


Figure 3: Dynamic NPBN with only DTR and three periods of observations.

Figure 3 shows the DNPBN used further in this analysis. The marginal distributions of the nodes remain unchanged when compared with the static models. Some of the rank correlations are the same as the ones in Figure 2, and the remaining are calculated as explained above. Some of the numbers shown on the arcs correspond to conditional rank correlations, which explains their very small (almost zero) values.

4 Unintuitive results or incorrect intuitions?

Having built and quantified the models for our example application, we shall now investigate some of the probabilistic phenomena described in the introduction of this paper. The joint distributions represented by the NPBNs discussed in the previous sections clearly depart from joint normality, given that some of the univariate margins (shown in the histograms) are not normal. Most of our probabilistic intuitions are based on the behaviour of the joint normal distribution. Updating a joint distribution when new information becomes available equates to conditioning the joint distribution. Conditioning a non normal joint distribution may generate surprising results.

4.1 Negative learning

Negative learning happens when an unexpected piece of information increases the uncertainty. Inferring decadal trends on the basis of 10 years of data bears the full brunt of natural variability, as reflected in the low correlations showed on the left hand side of in Figure 2. In contrast, the right hand side of Figure 2 shows the state of play in 2050, after 30 years of observations. The increased correlations reflect the greater certainty in long term trends acquired after 30 years of observation. We start with the model from the right hand side of Figure 2 (and repeated on the left hand side of Figure 4 for clarity) and perform inference (conditionalizing) assuming a high (higher than expected) value of DTR was observed when using the old measurement system. The right hand side of Figure 4 shows the results of such a measured value (0.5). Note that the mean value for ECS has risen to 5.83, and uncertainty (standard deviation) for ECS has increased from 1.73 to 2.09. The conditional distributions are now shown in black and the grey histograms represent the unconditional margins, provided for comparison. The unexpectedly high measured value caused the mean value and also the uncertainty in ECS to rise. This of course is impossible under the simple normal error model. Notice that the distributions of ECS and decadal temperature rise are far from normal, so a standard normal model is inappropriate.

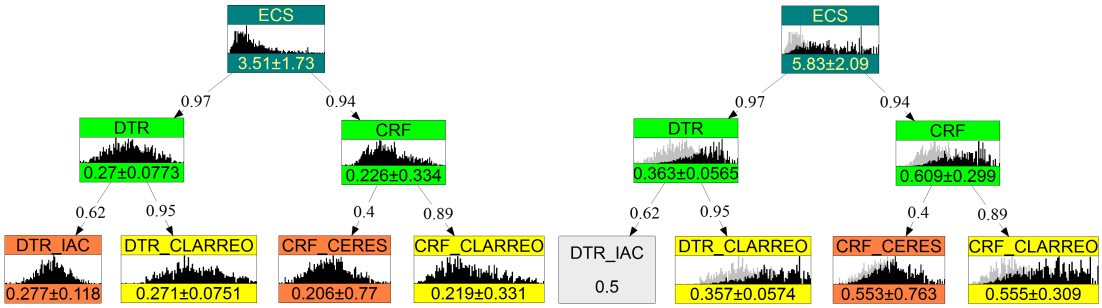


Figure 4: NPBN after 30 years of observing decadal trends that follow the launch of CLARREO in 2020.

Intuitively, noisy measurements and large uncertainties are the playground for negative learning, which represent a reason for quantifying the full breadth of uncertainty in our models. If our measurements are independent of the quantity of interest (described by zero correlation), then the measurement has no effect, i.e. the conditional distribution of the quantity of interest post observation is the same as pre-observation. On the other hand, if the measurement has a very small error, which is described by a correlation close to one, then the measurement eliminates almost all uncertainty in the quantity of interest. Whatever the measured value, the conditional variance of the quantity of interest will be nearly zero. Between these extremes there are cases, depending on the unconditional distributions, where an extreme measurement can drag away part, but not most of the probability mass, causing higher conditional variance. In this example, negative learning is more difficult in 2050 than it is in 2030, and all but impossible in 2080.

We study now the effects of negative learning using the DNPBN in Figure 3. Negative learning diminishes when observations are accumulated in time. The diminishing process is faster when the new measuring system is used instead of the current one. Figure 5 shows the conditional standard deviation of the climate sensitivity when every 5 years we observe extreme values of the DTR. The solid line corresponds to the current (IAC) system and the dashed line corresponds to the new (CLARREO) measuring system. To obtain this picture, another DPBN model (presented in the Appendix), similar to the one in Figure 3, was built, quantified and conditionalised. For this DNPBN, 11 time steps were considered, every 5 years

starting with 2030.

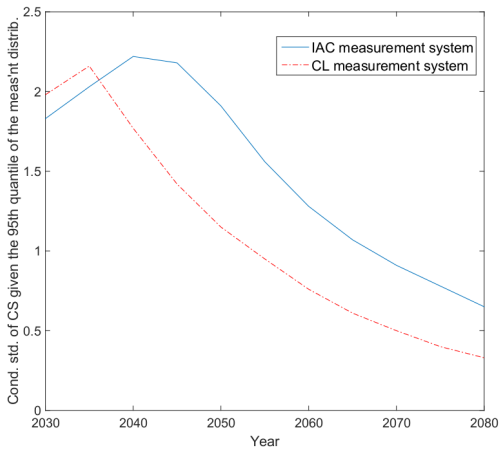


Figure 5: Conditional standard deviation of ECS given extreme values of DTR for the current (IAC) and enhanced (CL) measuring system.

Given the complexity of the problem, the simplifications made by the models and the intricacies of the uncertainties, it is tempting to think that slightly different correlations may not lead to the negative learning phenomenon observed above. The fact that this phenomenon is difficult to observe in 2050, and impossible in 2080 is another incentive for a sensitivity analysis that verifies this conjecture. The dependence between the variables of these models (as measured by the rank correlations) changes with time. The links become stronger. When several concurrent changes of the rank correlation values (within the same model) of order 10^{-1} are made, and the same inferences are performed, negative learning is still observed.

How much should the weakest correlation change for the negative learning to subside? We will use the NPBN from the left hand side of Figure 2 to further explore how the magnitude of the rank correlation between DTR and DTR_IAC can influence negative learning. The figure below illustrates this behaviour.
Note that when the correlation achieves a value of almost 0.63, the conditional variance of the climate sensitivity starts to decrease, i.e. the uncertainty decreases when new information is assimilated into the system. It is interesting to notice that in the 2050 model, the correlation between DTR and DTR_IAC is 0.62. However, negative learning is still observed in this model due to all the other changed parameters (see Figure 4).

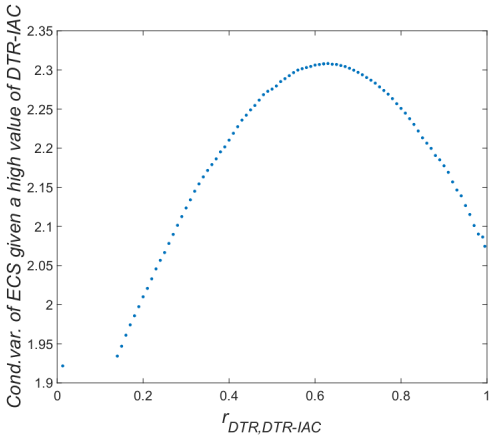
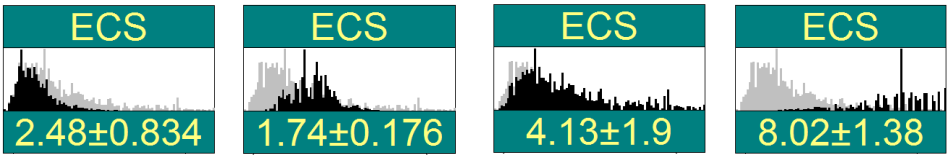


Figure 6: Conditional variance of climate sensitivity (ECS) given a high value of DTR_IAC for various rank correlations between DTR and DTR_IAC.

4.2 Discordant agreement

Let us now investigate different measurements of the same system, returning the same values. Intuition dictates that such measurements should yield the same conclusions. We consider the NPBN from the left hand side of Figure 4, representing the year 2050, 30 years after CLARREO was launched. Assume temperature is observed to have a very small value, 0.15, and this observation is consistent between the two measurement systems. Conditioning the model on $DTR\_IAC = 0.15$ changes the distribution of ECS as in Figure 7a. Both the expectation and the standard deviation decrease considerably. However, the decrease is even more dramatic if the same value of DTR is observed using the new measurement system. The conditional distribution of ECS given DTR_CLARREO =0.15 is shown in Figure 7b. Not only the values of the mean and the standard deviation change, but the shape of the distribution changes as well.

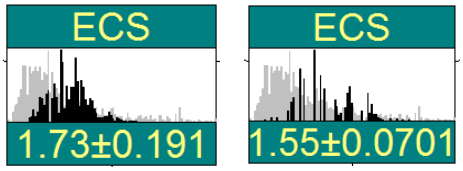
Big differences are also observed when looking at the conditional distribution of ECS given an extreme (this time large) value of CRF, as measured by the two different systems.This is shown in Figures 7c and 7d. An increase of both the mean and the standard deviation of the ECS are expected. However, what is not expected is such a discrepancy between conditioning on the current system (4.13 ± 1.9) and the new system (8.02 ± 1.38). The conditional expectation of ECS almost doubles when the same value for CRF is observed using CLARREO.



(a) DTR_IAC = 0.15 **(b)** DTR_CL = 0.15 **(c)** CRF_CERES = 1 **(d)** CRF_CL = 1

Figure 7: Conditional distribution of ECS given the same observations, measured on different measurement systems, using the static NPBN model from 2050.

Using the DNPBN from Figure 3 to investigate the same phenomenon, we condition on the DTR value used above, 0.15, first in the IAC system, in all of the three time steps, i.e. DTR_IAC30 = DTR_IAC50 = DTR_IAC70 = 0.15. The conditional distribution of ECS given the 3 conditioned variables is presented in Figure 8a. When conditioning on DTR_CLARREO30 = DTR_CLARREO50 = DTR_CLARREO70 = 0.15 the conditional distribution of ECS changes by what may seem as little. This new conditional distribution is shown in Figure 8b. The expected value of the ECS's conditional distribution when the current measurement system is used is larger than the 99th quantile of the conditional distribution obtained when using the new system.



(a) DTR_IAC = 0.15 **(b)** DTR_CL = 0.15

Figure 8: Conditional distribution of ECS given the same observations, measured on different measurement systems, using the DNPBN model.

In the cases investigated above, the same measurements generate quite different conclusions. The different uncertainties in the measurements' errors affect the mean in ways different than those the simple error model trained our intuitions with.

4.3 Obsolescence in time

Would maintaining older measuring systems still be useful even after better systems have become available? One can argue that the older systems can still provide useful information, if not on their own, at least when used in conjunction with the new

systems.

Let us assume that in the 2050 (static) model measurements conducted with the existing systems for DTR and CRF found measured values equal to expected values. Such measurements would yield $ECS = 3.33 \pm 1.27$. Conditioning the CLARREO measured DTR and CRF at their expected values yields $ECS = 3.18 \pm 0.51$. This is a substantial uncertainty reduction relative to the existing systems. However, the latter values would scarcely change if the older noisier systems also returned their expected values ($ECS = 3.17 \pm 0.5$). In this case the older system adds little to the knowledge gained by the newer system.

If, along with conditioning the enhanced systems measurements on their expected value, we also condition the older systems measurements to be extreme, that is $DTR\_IAC = 0.5$ and $CRF\_CERES = 2$, we would find $ECS = 3.5 \pm 0.59$ (see Figure 9). Apparently the older systems can influence our results, even after the introduction of the much more accurate enhanced EOS systems. However their contribution to the knowledge about ECS depends on the actual measured values. While older systems may indeed be uninformative if everything happens as we expect, they can add value when things do not turn out as we expect; categorical judgements cannot be sustained. This conclusion would not be anticipated on the basis of the simple normal error model.

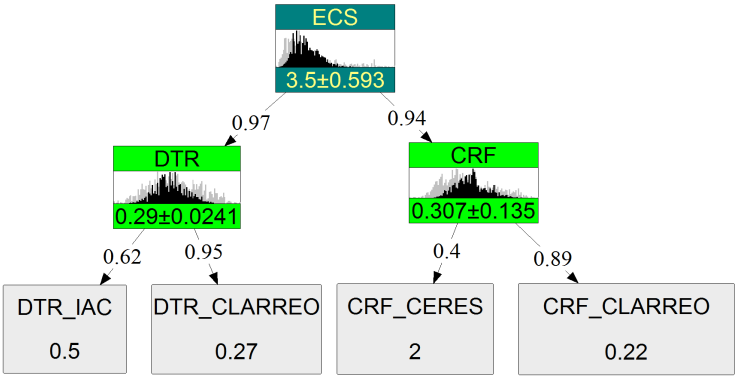


Figure 9: Conditioning the enhanced measuring system variables at their expected values followed by conditioning the current measuring systems' variables at their expected values in the static 2050 model.

Given the two measuring systems for global temperature rise in our DNPBN model, we are now interested in investigating the obsolescence phenomenon within the

DNPBN. Expectedly, when conditioning in time on expected values on both systems, the current one becomes obsolete. Nonetheless, when the new measuring system is conditioned on the expected value of their measurements and the current IAC system on the extreme 0.5 value, then the conditional distribution of ECS is different than the one obtained by only conditioning on the new measuring system on its expected values. Figures 10 and 11 provide this comparison.

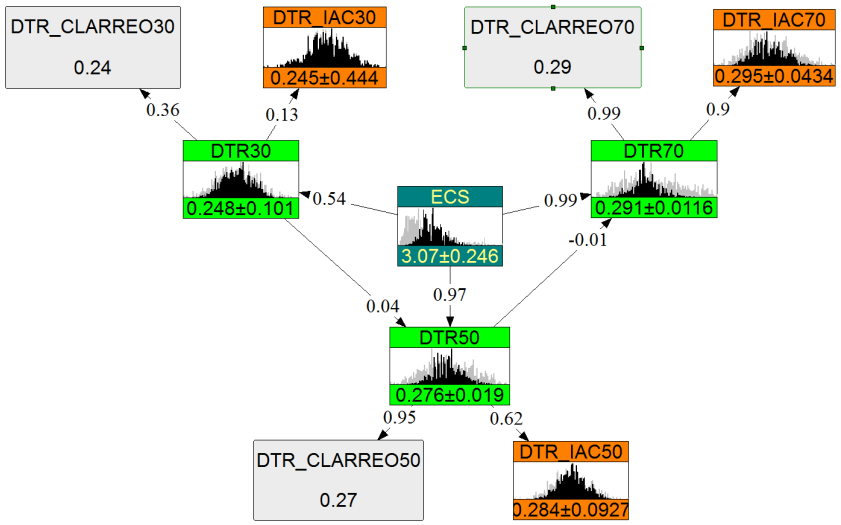


Figure 10: DNPBN when conditioning on the expected values for CLARREO measurements at every time step.

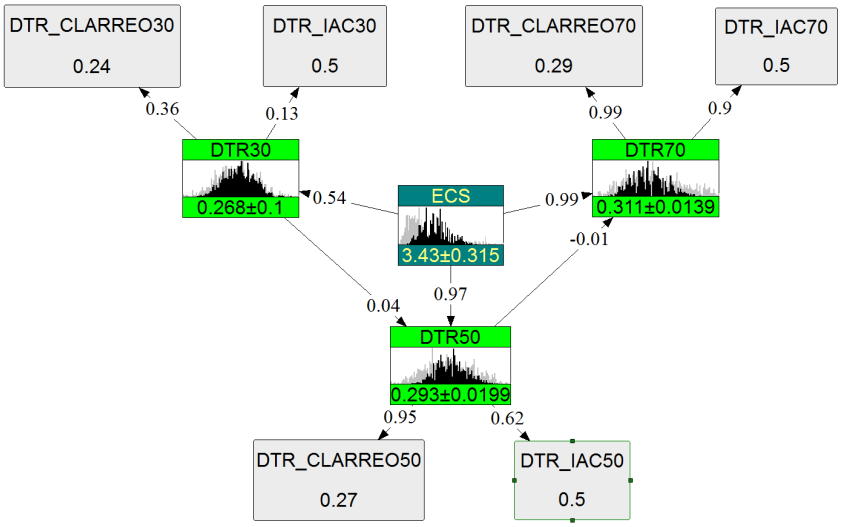


Figure 11: DNPBN when conditioning on the expected values for CLARREO measurements and the 0.5 value for the IAC measurements, at every time step.

The ECS conditional distributions showed in Figures 10 and 11 have different pa-

rameters and quite different shapes. Hence, they are found to be different when performing a two-sample Kolmogorov-Smirnov goodness-of-fit hypothesis test. The two conditional distributions are plotted on top of one another in form of histograms. Figure 12 allows us to observe their differences better.

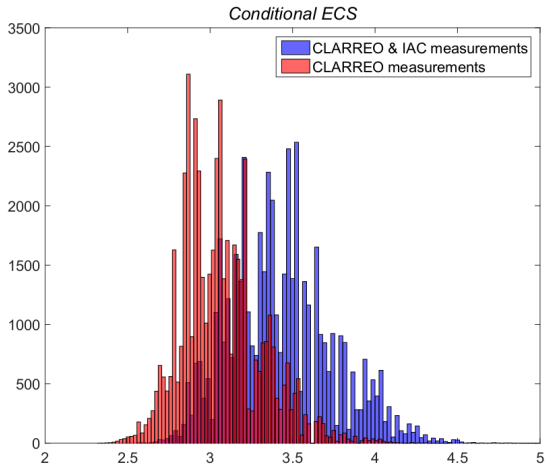


Figure 12: Conditional ECS – approximated densities.

5 Conclusions

Understanding complex probabilistic models under the climate change paradigm unveils great challenges. The drawbacks of *classical* modes of reasoning based on simple statistical models, are highlighted by our analysis. New observations do not always reduce uncertainty and older measuring systems may sometimes supply useful information. Independent errors are always the easiest choice, but correlated errors are more realistic. They can easily be modelled using NPBNs. Moreover, the dynamic nature of the models presented here can be incorporated into DNPBNs. Observing unintuitive effects of a more sophisticated approach should not be used as a reason for avoiding a more rigorous, realistic model.

We used BNs to investigate and support our claims. BNs with variables taking continuous one dimensional marginal distributions and rank correlations lend themselves for representing inferences from noisy data. BNs effortlessly perform operations of computing conditional distributions that would be extremely onerous analytically. The type of BNs used in this research, the NPBNs have easily

handled models with over a thousand variables (Hanea et al. 2015). Furthermore, DNPBNs model temporal influences, which acknowledge the dynamical nature of the problem.

References

Ale, B., L. Bellamy, J. Cooper, R. Cooke, L. Goossens, D. Kurowicka, O. M. Napoles, A. Roelen, and J. Spouge (2009). Further development of a Causal model for Air Transport Safety (CATS); Building the mathematical heart. *Reliability Engineering and Safety Systems 94(9)*, 1433–1441.

Cooke, R. (1991). *Experts in Uncertainty : Opinion and Subjective Probability in Science*. Environmental Ethics and Science Policy Series. Oxford University Press.

Cooke, R. M., A. Golub, B. A. Wielicki, D. F. Young, M. G. Mlynczak, and R. R. Baize (2016). Using the social cost of carbon to value earth observing systems. *Climate Policy*.

Cooke, R. M. and B. A. Wielicki (2016a). Bayesian nets for combining disparate information sources. *Resources for Future Report report*.

Cooke, R. M. and B. A. Wielicki (2016b). Probabilistic thinking and the art of constructive disagreement. *Resources for Future Report*.

Cooke, R. M., B. A. Wielicki, D. A. Young, and M. G. Mlynczak (2014). Value of information for climate observing systems. *Environ Syst Decis 34*, 98–109.

Hanea, A., D. Kurowicka, and R. Cooke (2006). Hybrid Method for Quantifying and Analyzing Bayesian Belief Nets. *Quality and Reliability Engineering International 22*(6), 613–729.

Hanea, A., D. Kurowicka, R. Cooke, and D. Ababei (2010). Mining and visualising ordinal data with non-parametric continuous BBNs. *Computational Statistics and Data Analysis 54*(3), 668–687.

Hanea, A., O. Morales-Napoles, and D. Ababei (2015). Non-parametric Bayesian networks: Improving theory and reviewing applications. *Reliability Engineering and System Safety 144*, 265284.

Hanea, A. M., O. M. Napoles, and D. Ababei (2015). Non-Parametric Bayesian Networks: Improving Theory and Reviewing Applications. *Reliability Engineering and Safety Systems 144*, 265–284.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Kurowicka, D. and R. Cooke (2004). Distribution - Free Continuous Bayesian Belief Nets. Proceedings Mathematical Methods in Reliability Conference.

Kurowicka, D. and R. Cooke (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley.

Leroy, S. S., J. G. Anderson, and G. Ohring (2008). Climate signal detection times and constraints on climate benchmark accuracy requirements. *Journal of Climate 21*, 841–846.

Morales, O., D. Kurowicka, and A. Roelen (2008). Eliciting conditional and unconditional rank correlations from conditional probabilities. *Reliability Engineering & System Safety 93*(5), 699 – 710.

Nelsen, R. (1999). *An Introduction to Copulas*. Lecture Notes in Statistics. New York: Springer - Verlag.

Nordhaus, W. and P. Sztorc (2013). *DICE 2013R: Introduction and user's manual (2nd ed.)*.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo: Morgan Kaufman Publishers.

Roe, G. H. and M. B. Baker (2007). Why is climate sensitivity so unpredictable? *Science 318(5850)*, 629–632.

Wielicki, B. A., D. F. Young, M. G. Mlynczak, K. J. Thome, S. Leroy, J. Corliss, J. G. Anderson, C. O. Ao, R. Bantges, F. Best, K. Bowman, H. Brindley, J. J. Butler, W. Collins, J. A. Dykema, D. R. Doelling, D. R. Feldman, N. Fox, X. Huang, R. Holz, Y. Huang, Z. Jin, D. Jennings, D. G. Johnson, K. Jucks, S. Kato, D. B. Kirk-Davidoff, R. Knuteson, G. Kopp, D. P. Kratz, X. Liu, C. Lukashin, A. J. Mannucci, N. Phojanamongkolkij, P. Pilewskie, V. Ramaswamy, H. Revercomb, J. Rice, Y. Roberts, C. M. Roithmayr, F. Rose,

S. Sandford, E. L. Shirley, W. L. S. Sr., B. Soden, P. W. Speth, W. Sun, P. C. Taylor, D. Tobin, and X. Xiong (2013). Achieving climate change absolute accuracy in orbit. *Bulletin of the American Meteorological Society 94*, 1519–1539.

Appendix

The first part of the appendix will collate a number of definitions and concepts that may prove useful. All of them are necessary for establishing the probabilistic background of the NPBN framework, and can be found in Kurowicka and Cooke (2006) together with the original references.

Definition 1. *The* copula *of two continuous random variables X and Y is the joint distribution of $F_X(X)$ and $F_Y(Y)$, where F_X, F_Y are the cumulative distribution functions of X, Y respectively. The copula of (X, Y) is a distribution on $[0,1]^2 = \boldsymbol{I}^2$ with uniform marginal distributions.*

An overview of copulae can be found in Joe (1997) and Nelsen (1999). Here, we only give the definition of the *normal copula*, which is the only copula used in the analysis from this paper. The normal copula is a one parameter copula.

Definition 2 (Normal copula)**.** *If Φ_ρ is the bivariate normal cumulative distribution function with product moment correlation ρ and Φ^{-1} the inverse of the standard univariate normal distribution function then, the normal copula is defined as:*

$$C_\rho(u, v) = \Phi_\rho\left(\Phi^{-1}(u), \Phi^{-1}(v)\right), \quad (u, v) \in \boldsymbol{I}^2.$$

The joint normal distribution may be described by having normal marginal distributions and a normal copula. This means that any continuous non-normal joint distribution with its dependence realised by a normal copula could be transformed to a join normal distribution by simply transforming the margins to normals.

One parameter copulae in general may be parametrised by the rank correlation, which is different than the linear (product moment) correlation ρ. For the normal

copula the relationship between the product moment correlation (ρ) and the rank correlation (r), is given by the Pearson's transformation.

Proposition 1. *Let (X, Y) be a random vector with the joint normal distribution, then:*

$$\rho(X, Y) = 2\sin(\tfrac{\pi}{6} \cdot r(X, Y)).$$

The normal copula inherits many properties from the joint normal distribution, amongst which the fact that the partial correlations are equal to the conditional correlations. Partial correlations can be defined in terms of partial regression coefficients. Let us consider variables X_i with zero mean and standard deviations σ_i, $i = 1, ..., n$. Let the numbers $b_{12;3,...,n}$,...,$b_{1n;2,...,n-1}$ minimise:

$$E\left((X_1 - b_{12;3,...,n}X_2 - ... - b_{1n;2,...,n-1}Xn)^2\right).$$

Definition 3. *The partial correlation of X_1 and X_2 based on X_3,..., X_n is:*

$$\rho_{12;3,...,n} = sgn(b_{12;3,...,n})(b_{12;3,...,n}b_{21;3,...,n})^{\frac{1}{2}}.$$

Equivalently we could define the partial correlation as:

$$\rho_{12;3,...,n} = -\frac{C_{12}}{\sqrt{C_{11}C_{22}}},$$

where C_{ij} denotes the $(i, j)^{th}$ cofactor of the correlation matrix.

The partial correlation $\rho_{12;3,...,n}$ can be interpreted as the correlation between the orthogonal projections of X_1 and X_2 on the plane orthogonal to the space spanned by X_3,...,X_n.

Partial correlations can be computed from correlations with the following recursive formula:

$$\rho_{12;3,...,n} = \frac{\rho_{12;4,...,n} - \rho_{13;4,...,n} \cdot \rho_{23;4,...,n}}{((1 - \rho^2_{13;4,...,n}) \cdot (1 - \rho^2_{23;4,...,n}))^{\frac{1}{2}}}.$$

In the last part of this appendix we will show the DNPBN used to obtain Figure 5.

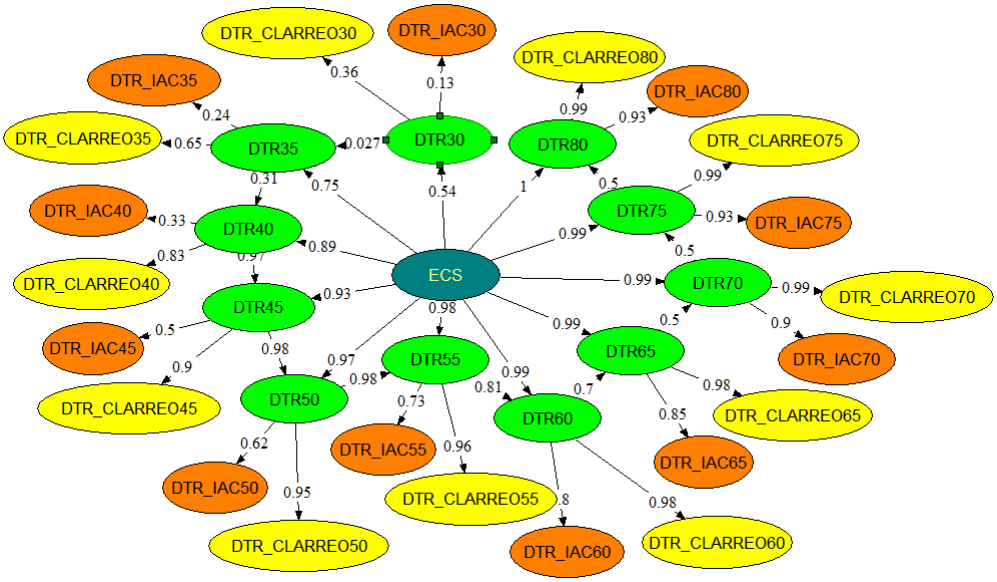


Figure 13: DNPBN with 11 time steps, every 5 years starting with 2030.