



## **Análise e Previsão de Padrões de Consumo com Modelo Escondido de Markov**

**FRANCISCO MANUEL PEREIRA RESENDE DE ANDRADE**

novembro de 2021

# ANÁLISE E PREVISÃO DE PADRÕES DE CONSUMO COM MODELO ESCONDIDO DE MARKOV

Francisco Manuel Pereira Resende de Andrade



Departamento de Engenharia Eletrotécnica

Mestrado em Engenharia Eletrotécnica – Sistemas Elétricos de Energia

**2021**



Relatório elaborado para satisfação parcial dos requisitos da Unidade Curricular de TEDSE  
– Dissertação/Estágio do Mestrado em Engenharia Eletrotécnica – Sistemas Elétricos de  
Energia

Candidato: Francisco Manuel Pereira Resende de Andrade, Nº 1161568,  
1161568@isep.ipp.pt

Orientação científica: Zita Vale, zav@isep.ipp.pt; Tiago Manuel Campelos Ferreira Pinto,  
tcp@isep.ipp.pt



Departamento de Engenharia Eletrotécnica  
Mestrado em Engenharia Eletrotécnica – Sistemas Elétricos de Energia

**2021**



Para os meus pais, que sempre acreditaram em mim.



## *Agradecimentos*

A realização deste projeto só foi possível graças à ajuda e apoios que tive desde o início.

Aos Professores Tiago Pinto e Zita Vale, pela orientação, disponibilidade e ajuda ao longo de toda a dissertação.

À minha família por todo o apoio e confiança que me deram durante a minha vida.

Aos meus amigos por se manterem ao meu lado e por toda a ajuda que me deram. Um especial obrigado à Filipa, Leonor, Sofia e Diva, porque sem vocês não seria o que sou hoje.



## *Resumo*

O consumo de energia nos edifícios é significativo. De forma a acompanhar a demanda de energia, começou-se a recorrer a energias renováveis para produzir mais energia de uma forma sustentável. Isso levou, no entanto, a um aumento da penetração de energia de natureza variável na rede. Para que se possa fazer face à variação do lado da geração, é necessário o desenvolvimento de modelos de gestão e adaptação do lado do consumo. Uma forma de desenvolver métodos para diminuir e gerir o consumo em habitações é conseguir prever o consumo das habitações. Há muitas metodologias que podem ser usados para identificar e prever dados, mas para esta dissertação foi usado o *Hidden Markov Model (HMM)*, uma vez que muitos artigos demonstraram bons resultados na sua aplicação a problemas com características semelhantes. Foi usado a livreria “HMM” existente no programa R. Dos testes efetuados com o HMM, foi possível verificar que este consegue prever os consumos de energia. No entanto, em diferentes testes de outros autores, os seus HMMs obtiveram melhores resultados. Uma limitação encontrada foi a falta de possibilidade de personalização do programa utilizado, o que leva a que em situações mais complexas, os resultados obtidos podem não ser satisfatórios. Ao efetuar o último teste, foi feito um treino com o HMM de vários dias, para de seguida prever os estados nos dias seguintes. Quando a previsão foi efetuada, comparou-se esses dias previstos com os dados reais existentes na base de dados, tendo conseguido prever corretamente 48% dos estados escondidos e 72% dos estados observáveis.

### *Palavras-Chave*

Consumo de eletricidade em edifícios residenciais, previsão de consumo, *Hidden Markov Model*, R.



## *Abstract*

The energy consumption on household buildings is high. To keep up with the demand, it started to use renewable energies to produce more energy in a sustainable way, but it led to the increase of the energy penetration in the grid. To face the variation on the generation side, it is necessary to develop methodologies to predict the energy consumption of the habitants. There is a big number of different methodologies that can identify and predict the consumption, but for this dissertation was used the Hidden Markov Model, since a good number of scientific articles showed a good outcome when applied to problems with similar characteristics. It was used the “HMM” library in the program R. It was possible to testify that the HMM can predict the consumption. However, other authors tests were able to get better results with their HMMs. One limitation is the lack of customization of the used program, so it can lead to worst results if the problem is more complex. With the last test, was possible to train the HMM for multiple days and predict the consumption for the next three days. When the prediction was done, the data was compared to the real data and the HMM was able to predict correctly 48% of the hidden states and 72% of the observable states.

### ***Keywords***

Household energy consumption, Consumption prediction, Hidden Markov model, R.



# Índice

<b>AGRADECIMENTOS</b> .....	<b>I</b>
<b>RESUMO</b> .....	<b>III</b>
<b>ABSTRACT</b> .....	<b>V</b>
<b>ÍNDICE</b> .....	<b>VII</b>
<b>ÍNDICE DE FIGURAS</b> .....	<b>IX</b>
<b>ÍNDICE DE TABELAS</b> .....	<b>XIII</b>
<b>ACRÓNIMOS</b> .....	<b>XV</b>
<b>1. INTRODUÇÃO</b> .....	<b>1</b>
1.1.CONTEXTUALIZAÇÃO .....	1
1.2.OBJETIVOS .....	4
1.3.CALENDARIZAÇÃO .....	5
1.4.ORGANIZAÇÃO DO DOCUMENTO.....	5
<b>2. ESTADO DA ARTE</b> .....	<b>7</b>
2.1.ANÁLISE DE COMPONENTES PRINCIPAIS .....	9
2.2.MAPA AUTO-ORGANIZADOS.....	10
2.3.PERCEPTRON MULTICAMADAS.....	11
2.4.AGRUPAMENTO HIERÁRQUICO.....	11
2.5.MOTIF .....	11
2.6.AGRUPAMENTO K-MEANS .....	12
2.7.REDE NEURONAL ARTIFICIAL.....	13
2.8.MODELO HÍBRIDO .....	13
2.9.MODELOS ESCONDIDOS DE MARKOV .....	13
2.10.DISSCUSSÃO .....	16
<b>3. MODELOS ESCONDIDOS DE MARKOV</b> .....	<b>19</b>
3.1.CONSTITUIÇÃO DO HMM .....	21
3.2.PROBLEMA DE PROBABILIDADE .....	22
3.3.DECODIFICAÇÃO DO HMM .....	23
3.4.TREINO DO HMM.....	25
3.5.MODELO COMPUTACIONAL .....	25
<b>4. CASO DE ESTUDO</b> .....	<b>27</b>
4.1.LEVANTAMENTO E TRATAMENTO DOS DADOS .....	27
4.2.TESTE 1 – 4 ESTADOS ESCONDIDOS E 9 ESTADOS OBSERVÁVEIS .....	29

4.3. TESTE 2 – 4 ESTADOS ESCONDIDOS A 25% E 9 ESTADOS OBSERVÁVEIS .....	35
4.4. TESTE 3 – 2 ESTADOS ESCONDIDOS E 2 ESTADOS OBSERVÁVEIS .....	38
4.5. TESTE 4 – 2 ESTADOS ESCONDIDOS E 2 ESTADOS OBSERVÁVEIS COM MENOS DADOS .....	40
4.6. TESTE 5 – 2 ESTADOS ESCONDIDOS E 2 ESTADOS OBSERVÁVEIS COM PADRÃO DA TELEVISÃO .....	43
4.7. DISCUSSÃO DOS RESULTADOS .....	46
<b>5. CONCLUSÕES.....</b>	<b>49</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>51</b>

## Índice de Figuras

Figura 1 – Exemplo de HMM	21
Figura 2 - Evolução temporal dos equipamentos num dia	29
Figura 3 - Percentagem de cada estado escondido para o teste 1	30
Figura 4 - Probabilidades de transição para o teste 1 (estado escondido - estado escondido)	31
Figura 5 - Probabilidades de emissão para o teste 1 (estado escondido – estado observável)	32
Figura 6 - Estado 1 do teste 1	32
Figura 7 - Resultados do teste 1 com <i>Viterbi</i>	33
Figura 8 - Comparação dos estados escondidos e observáveis no teste 1	34
Figura 9 - Resultados do teste 1 com <i>Baum-Welch</i>	34
Figura 10 - Resultados do teste 1 para 70/30 com <i>Viterbi</i>	35
Figura 11 - Número de cada estado do teste 2	36
Figura 12 - Probabilidade de transição do teste 2	36
Figura 13 - Probabilidade de emissão para teste 2	37
Figura 14 - Resultados do teste 2	37
Figura 15 - Número dos estados escondidos e observáveis do teste 3	38
Figura 16 - Probabilidade de transição para teste 3	39
Figura 17 - Probabilidade de emissão para teste 3	39

Figura 18 - Resultados do teste 3	39
Figura 19 - Comparação dos estados escondidos e observáveis no teste 3	40
Figura 20 - Número dos estados escondidos e observáveis do teste 4	41
Figura 21 - Probabilidade de transição para teste 4	41
Figura 22 - Probabilidade de emissão para teste 4	41
Figura 23 - Resultados do teste 4	42
Figura 24 - Comparação dos estados escondidos e estados observáveis no teste 4	42
Figura 25 - Número dos estados escondidos e observáveis do teste 5	44
Figura 26 - Probabilidade de transição para teste 5	44
Figura 27 - Probabilidade de emissão para teste 5	44
Figura 28 - Resultados do teste 5	45
Figura 29 - Comparação dos estados escondidos e observáveis no teste 5	46





## *Índice de Tabelas*

Tabela 1 - Calendarização do projeto	5
Tabela 2 - Vantagens e Desvantagens dos Métodos/Algoritmos	16
Tabela 3 - Resumo dos resultados	46



## *Acrónimos*

- ANN – Artificial Neural Network
- DSM – Demand-Side Management
- FCM – Fuzzy M-Means
- HMM – Hidden Markov Model
- MLP – Multilayer Perceptrons
- PCA – Principal Component Analyses
- SOM – Self-Organizing Maps



# 1. INTRODUÇÃO

O presente trabalho tem como objetivo realizar uma previsão do consumo de energia para edifícios residenciais.

## 1.1. CONTEXTUALIZAÇÃO

O consumo de energia tem um valor considerável nos edifícios residenciais, que tem acontecido devido ao aumento da população, aumento dos salários e a procura de maior conforto. Atualmente, 29 % da energia consumida globalmente é em habitações residenciais, o que significa que pode haver muito potencial para realizar poupanças na energia (Abu Al-Haija, 2021; Mokhov et al., 2021).

Uma das principais formas de produzir energia, costumava ser a partir de combustíveis fósseis, mas com o aumento da demanda de energia, foi necessário recorrer a outras fontes, uma vez que fontes não renováveis são limitadas e acompanhadas de grandes libertações de gases com efeito de estufa (Hossain, 2012).

Nos últimos anos, o uso energias renováveis têm aumentado. Essas energias são fontes de energia limpas, onde existe diversos tipos, como solar, hídrica e eólica, que podem ser utilizadas sempre que houver disponibilidade (Hossain, 2012). No entanto, com o aumento da penetração dessas energias, concluiu-se que é necessário desenvolver novas tecnologias (Flores-Quiroz et al., 2016).

Com o aumento dos preços de eletricidade, muitos consumidores são levados à produção da própria energia, como é o caso do exemplo de Fett, Keles, Kaschub e Fichtner (2019), com os painéis fotovoltaicos.

Como foi referido anteriormente, o consumo mundial em edifícios residenciais, é aproximadamente de 29% da energia total consumida. Por isso, é muito importante diminuir o seu consumo. Para o fazer, Sütterlin, Brunner e Siegrist (2011) referiram que é necessário identificar e descrever os diferentes tipos de consumidores, para se saber quais serão as melhores estratégias para aplicar para cada tipo deles.

O uso de diferentes tarifas no mercado de eletricidade tem levado à mudança do consumo e do comportamento dos consumidores. Rosin, Hõimoja, Möller e Lehtla (2010) informam que é preciso ter em mente as necessidades dos consumidores, como os preços, hábitos, conveniência, disponibilidade e qualidade. Para se poder fornecer as melhores condições aos consumidores é essencial analisar os seus padrões. Os autores consideraram que é muito difícil conseguir reduzir os custos para o consumidor se estes não tiverem nenhum gerador de energia, então consideram trocar cargas ou substituir cargas menos eficientes por umas mais eficientes. Existem casos em que é possível mudar os consumos de um horário para o outro, como é o caso de máquinas de lavar roupa (Pinto et al., 2018; Santos et al., 2012)

Um estudo realizado por Iliopoulos, Esteban e Kudo (2020) sugeriu que é possível transformar os consumidores residenciais de energia passivos em ativos, *prosumers*, de forma a criar uma rede inteligente, *smart grid*, em que há duas direções da eletricidade e informação, em vez que uma rede tradicional, que tem apenas uma direcção. O fator mais importante para se conseguir desenvolver projetos para redes inteligente, são os consumidores, pois são estes que podem contribuir para reservas de energia, alterações de consumo e aplicações amigas do ambiente. Os consumidores também podem participar como *prosumers* ao injetar energia na rede, por via de energias renováveis, como turbinas, painéis fotovoltaicos e sistemas de armazenamento de energia. Para realizar uma possível mudança para uma rede inteligente, é importante os consumidores aceitarem as novas tecnologias e mudar os seus comportamentos. Foi notado que as casas com baixos rendimentos são os que menos se interessam com estas mudanças. Uma grande parte do seu rendimento é gasto em eletricidade e se essas casas participassem na rede inteligente, poderiam poupar mais dinheiro. Os autores puderam concluir após o estudo realizado, que

uma forma de atrair as casas com baixos rendimentos é uma potencial redução dos gastos em eletricidade. No entanto, existe uma barreira que são os gastos iniciais em tecnologias, que podem ser ultrapassados ao educar as pessoas sobre os custos e benefícios das tecnologias de forma a combater os seus receios. Também se verificou que as casas com baixos rendimentos não se sentem atraídas para redes inteligentes por causa do ambiente. No estudo efetuado, foi possível averiguar que tendem a aderir mais às redes inteligentes se forem divulgadas informações relativas aos consumos e geração de energia pelos seus vizinhos, uma vez que cria uma “competição” para verem quem consegue produzir mais e consumir menos energia.

Um outro estudo efetuado por Lim, Lim, Wee, Li e Wang (2020) também comentou a distinção entre a rede tradicional e na rede inteligente. Foi abordado a necessidade da participação dos consumidores para se poder implementar uma rede inteligente. Neste artigo, falam sobre o *Demand Side Management* (DSM) que realiza o planeamento, implementação e monitorização de modo a influenciar o uso da eletricidade pelos consumidores. O DSM pode ser implementado de duas formas: baseado no preço ou baseado por incentivo. Quando a implementação é baseada no preço, o DSM influencia os consumidores a consumir com base nos preços reais ou pelas previsões do mercado do dia seguinte, enquanto se for baseado por incentivos, o DSM influencia o consumidor a mudar os seus consumos para alturas fora de picos de consumo de energia. O principal motivo da existência do DSM é influenciar os consumidores a consumir energia em alturas onde a procura é menor. Uma outra parte da pesquisa efetuada pelos autores, foi a combinação do DSM com o armazenamento de energia e energias renováveis. Se for possível os consumidores armazenarem energia que foi excessivamente produzida, esta pode ser utilizada quando for preciso, não havendo a necessidade de alterar as rotinas dos consumidores.

Outras formas de reduzir o consumo de energia é a utilização de um preço de eletricidade por escalões, equipamentos com poupança de energia e um comportamento de conservação de energia. O preço da eletricidade por escalão é uma das maiores estratégias do lado financeiro, utilizado para diminuir o consumo nas habitações, mas não é a melhor forma de o fazer, uma vez que os clientes pensam que é uma maneira dos distribuidores de eletricidade aumentarem os preços da energia. A nível dos equipamentos de poupança de energia, é um dos métodos que as pessoas mais aderem, mas com o progresso tecnológico, não tem havido

uma grade promoção desses equipamentos. Por fim, um comportamento de conservação de energia é mais sustentável e durador a longo prazo (Mi et al., 2021).

Por causa de todos os fatores expostos, é necessário atuar para ajustar os preços de eletricidade e não sobrecarregar a rede. Uma das maneiras possíveis é a identificação, análise e previsão de padrões de consumo (Figueiredo et al., 2005). Ao identificar os diferentes tipos de consumidores e posteriormente analisá-los, é possível desenvolver diferentes soluções para não sobrecarregar a rede e diminuir os preços, de forma a não prejudicar a qualidade de vida dos consumidores (Vale et al., 2021). O presente trabalho aborda uma análise, identificação e previsão de padrões de consumo.

## **1.2. OBJETIVOS**

Esta dissertação foca-se na análise, identificação e previsão de padrões de consumo, tendo como principal objetivo realizar um estudo e desenvolver um modelo que é capaz de analisar, identificar e prever diferentes padrões de diferentes tipos de consumidores.

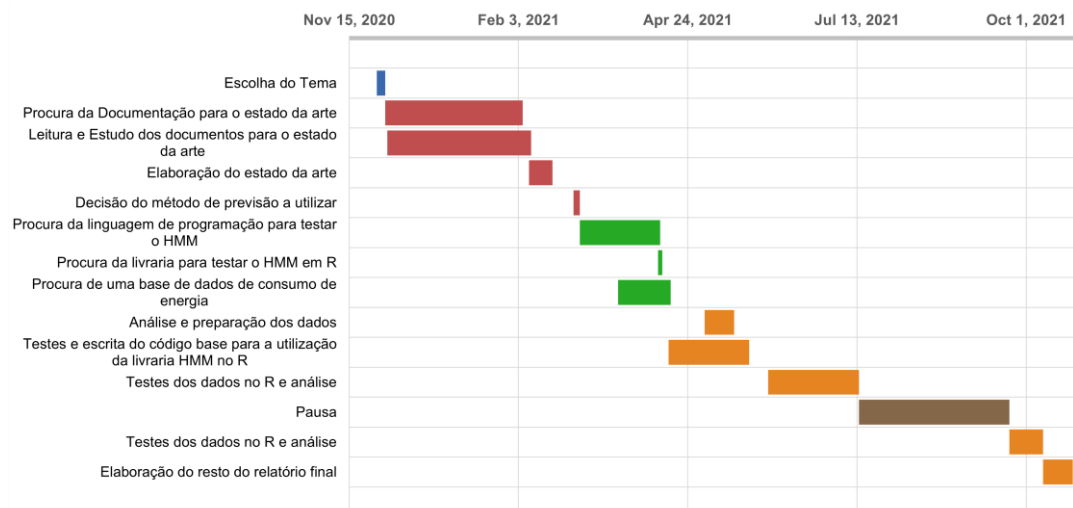
Os objetivos a atingir com esta dissertação são:

- Pesquisar modelos já existentes na literatura para a análise, identificação e previsão de padrões;
- Realizar um estudo do problema de identificação, análise e previsão de padrões de consumo de energia elétrica para diferentes tipos de consumidores;
- Desenvolver de metodologia capaz de analisar e identificar padrões de consumos, no qual faça uso de técnica de correlação e associação de dados, e de identificação de padrões;
- Fazer testes experimentais;
- Efetuar uma análise e avaliação dos resultados obtidos nos testes.

### 1.3. CALENDARIZAÇÃO

A calendarização deste trabalho de dissertação encontra-se na tabela 1.

Tabela 1 - Calendarização do projeto



Na calendarização está presente a escolha do tema. Para a fase de estudo, foi feita uma procura, leitura e estudo de documentos para o estado da arte, para depois na dissertação. De seguida, efetuou-se a decisão do método de previsão a utilizar. Com a decisão tomada, procurou-se a linguagem de programação e livreria que melhor se encaixava no problema atual. Também foi preciso procurar os dados para usar no HMM. Foram feitos testes e análise dos dados, terminando com a elaboração final do relatório.

### 1.4. ORGANIZAÇÃO DO DOCUMENTO

No Capítulo 1 é apresentada uma contextualização ao tema do presente trabalho, que explica os principais motivos que levaram à realização deste projeto, os principais objetivos a serem cumpridos e a calendarização do projeto. No capítulo seguinte, 2, está presente o levantamento do estado da arte do projeto, onde são explicados os métodos e algoritmos de identificação, análise e previsão mais utilizados, bem como os principais trabalhos efetuados com esses métodos, no tema do projeto. No seguinte capítulo, o 3º, é aprofundada a explicação dos HMMs. No 4ª capítulo são reunidos os principais testes efetuados para testar os HMMs desenvolvidos e é apresentada uma análise aos seus resultados. No 5º e último capítulo, está presente a conclusão onde é feito um resumo do relatório, uma discussão dos resultados obtidos e limitações.



## 2. ESTADO DA ARTE

O crescimento económico é um dos fatores que mais levou à alteração dos padrões de consumo de energia, uma vez que a população começou a ter mais poder de compra e assim podendo comprar mais equipamentos que requerem eletricidade para o seu funcionamento. Com isso, é essencial perceber como os padrões estão a evoluir com a expansão e crescimento económico, e a sua afetação no desempenho dos edifícios (Chai, 2018; Wei et al., 2018). É possível verificar que os padrões de consumo de energia não se diferenciam muito entre as pessoas cuja sua classe social é baixa ou média, mas variam imenso quando a classe social é elevada (Chai, 2018).

Atualmente o consumo de energia em edifícios está entre os 20% e 40% do total de consumo energético, e responsável por 33% das emissões de gases com efeito de estufa. Por causa disso, é importante conseguir identificar, analisar e prever padrões de consumo para melhorar a eficiência de energia, reduzir o impacto ambiental, melhorar a economia baseando na conservação de energia e energias renováveis (Wei et al., 2018).

As principais abordagens para se poder simular a energia consumida em edifícios e seguidamente identificar padrões, estão divididos em três categorias: abordagens baseadas na caixa branca, abordagens baseadas na caixa cinzenta e abordagens baseadas na caixa preta (Wei et al., 2018).

As abordagens baseadas numa caixa branca, são abordagens baseadas em física. No entanto, esta abordagem precisa de grande poder computacional (Wei et al., 2018). As abordagens baseadas na caixa cinzenta são feitas modificações à caixa branca e é adicionado métodos estatísticos, como por exemplo dados históricos do consumo de um edifício residencial. Para abordagem baseada na caixa preta são utilizados apenas os dados históricos e sem ter um conhecimento profundo do local físico. Nestes casos, a quantidade de dados utilizados é muito elevada e é principalmente utilizado algoritmos para a sua análise. Os modelos baseados na caixa preta são muitas vezes chamados como modelos baseados em dados.

Como o presente trabalho é focado na identificação, análise e previsão de padrões, é esperado ter um conjunto elevado de dados para a sua análise e assim conseguir-se identificar e prever novos padrões a partir dos dados já colecionados. Por esse motivo, a melhor abordagem para este tema é uma abordagem baseada na caixa preta. A maior parte dos métodos e algoritmos que irão requerer ser apresentados, irão requerer ter uma etapa de treino/aprendizagem, na qual são usados dados históricos (Jozi et al., 2019).

Muitos métodos para o estudo de padrões, recaem em diversas categorias, sendo algumas delas a redução, classificação, agrupamento (*clustering*) e regressão.

A redução é um passo inicial muito importante para o tratamento de dados de alta dimensão, pois é importante reduzir o tamanho dos dados, para se evitar problemas associados à grande dimensão de dados que possam surgir. Os dados iniciais muitas vezes não são apropriados para processar por causa da sua elevada dimensão e da sua redundância, que leva ao problema da grande dimensão, que é muito encontrado nos reconhecimentos de padrões, uma vez que implica que os números de dados para treino têm de aumentar exponencialmente com o número de características a encontrar, em ordem para o modelo ser mais preciso. Como normalmente não está disponível uma grande quantidade de dados, quanto mais características se pretender encontrar, mais impreciso o modelo começa a ser (Gutierrez-Osuna, 2002).

Os algoritmos para agrupamento de dados, *clustering*, têm como objetivo encontrar semelhanças ou uma relação espacial num conjunto de dados. Os métodos que envolvem este processo, têm de cumprir 3 passos:

1. Definir uma medida de distinção;
2. Definir um critério de agrupamento a ser otimizado;

3. Definir um algoritmo de busca para identificar possíveis agrupamentos, uma vez que pesquisar por todos os tipos possíveis de agrupamentos é inexecutável (Gutierrez-Osuna, 2002).

A regressão tem como objetivo prever um conjunto de propriedades. Estabelecer um modelo preditivo de um conjunto de variáveis independentes para outro conjunto de variáveis dependentes contínuas (Pinto et al., 2021). Para a previsão de padrões, pode ser usada a regressão no qual cada variável dependente é uma categoria (Gutierrez-Osuna, 2002).

## **2.1. ANÁLISE DE COMPONENTES PRINCIPAIS**

A Análise de Componentes Principais (PCA) é um método de redução. Tem principalmente dois objetivos: tratar dados de entrada de grandes dimensões e reduzi-los, de forma a criar uma representação de pequena dimensão e compreender as maiores tendências nos dados de entrada. O ponto principal no PCA é explorar a variação das características e a covariação entre elas (Burns, 2019).

PCA está sobre as suposições de Gaussian. As suposições de Gaussian são uma distribuição e são usados em diversas áreas, incluindo engenharia, estatística e física (Gutierrez-Osuna, 2002; Park et al., 2013).

A PCA é frequentemente usada em métodos estatísticos e tem aplicações em diversas áreas da estatística e de *machine learning*, que inclui *clusters*, redução de dimensão, reconhecimento facial, processamento de sinais, comparação de imagem, visualização e previsão. É capaz de reconhecer padrões em dados de grandes dimensões e ainda evidenciar semelhanças e diferenças nos dados de saída. Por outro lado, a PCA prioriza as principais tendências primeiro, o que nem sempre pode ser útil para o trabalho pretendido. Também complica detectar padrões com dados de elevada dimensão, uma vez que normalmente leva a interpretação simples (Burns, 2019).

No artigo de Abreu, Câmara Pereira, & Ferrão (2012), a PCA foi usada para dar uma perspectiva interna dos dados de padrões de consumo de energia e fazer uma análise de sensibilidade. Além disso, também procuraram comportamentos recorrentes que eram mantidos durante todo o ano numa base diária.

## 2.2. MAPA AUTO-ORGANIZADOS

Os mapas auto-organizados, SOM, têm como função detetar agrupamentos, *clusters*, no qual foi inspirado por auto organização e sistemas biológicos (Lobo, 2007; Miljković, 2017) e desenvolvido a partir de Redes Neurais Artificiais, ANNs (Wu et al., 2019). Com uma elevada dimensão de dados, o SOM representa esses dados em padrões com vetores de protótipo e podem ser vistos normalmente numa estrutura treliça de duas dimensões (Gutierrez-Osuna, 2002; Nikolaou et al., 2012). Os resultados são obtidos pela colaboração de três processos: competição, cooperação e adaptação (Gutierrez-Osuna, 2002). É necessário treinar este método, que é feito sem a necessidade de ser supervisionado e é capaz de classificar novas entradas em diferentes grupos com diferentes categorias de uma maneira neurobiológica (Miljković, 2017; Wei et al., 2018). Os padrões de entrada estão completamente ligados a todos os grupos por pesos adaptativos, que durante o treino, os padrões de entrada vizinhos são projetados numa treliça, correspondendo aos grupos adjacentes (Nikolaou et al., 2012). Tem uma aprendizagem competitiva de maneira a produzir uma representação discretizada de baixa dimensão e mapa de características da amostra de entrada.

As vantagens dos mapas auto-organizados são de serem capazes de teres as principais características dos dados inseridos e de conseguirem produzir uma representação discretizada e reduzida dos dados introduzidos de grandes dimensões. São capazes de ter uma aprendizagem sem supervisão e é usado em aplicações com redes de duas dimensões e retangulares, mas podem ser aplicados a redes hexagonais com três ou mais dimensões. As saídas deste método, podem ser imediatamente seguidas por outros métodos de classificação por agrupamento (Miljković, 2017).

No entanto, os mapas auto-organizados apresentam algumas desvantagens, uma vez que pode ser complicado o método funcionar com dados de muita elevada dimensão, uma vez que o custo computacional irá aumentar drasticamente (Gutierrez-Osuna, 2002; Wei et al., 2018). Uma outra desvantagem é se for introduzido, como dados de entrada, alguns dados não tratados, pois pode levar ao método sofrer oscilações (Wei et al., 2018).

### 2.3. PERCEPTRON MULTICAMADAS

Perceptron Multicamadas – *Multi-Layer Perceptron*, MLPs, é um tipo de redes neurais artificiais e é do tipo de redução e classificação. É uma rede *feed-forward* de elementos de processamento simples (neurónios), em que se ligam de uma maneira semelhante à de um circuito neural biológico. Cada neurónio faz uma soma ponderada das suas entradas e transforma-o, com uma função não linear, numa função sigmóide. É capaz de aprender regressões não lineares complexas, ajustando os pesos na rede por uma técnica de gradiente descendente conhecida como *back-propagation of errors*. O treino é efetuado em vários estágios e em cada estágio, o MLP processa as entradas de uma forma *feed-forward* e compara os resultados com os dados que se pretendia obter, propagando de volta esses erros para ajustar cada peso na rede de acordo com a sua contribuição para o erro geral. Além disso, a adição de camadas ocultas, pode permitir ao método ser mais eficiente, por utilizar menos pesos (Gutierrez-Osuna, 2002).

### 2.4. AGRUPAMENTO HIERÁRQUICO

Os agrupamentos hierárquicos, conhecidos como algoritmos aglomerativos, de acordo com a matriz de proximidades, organizam os dados numa estrutura hierárquica. O algoritmo começa por realizar uma partição inicial em *clusters singleton* até todos os elementos pertencerem ao mesmo grupo. Usando uma matriz de distância, é calculado todas as distâncias entre os elementos a serem agrupados. Em cada passo, dois elementos juntam-se, criando um nó. A partir daí, todos os nós seguintes são criados pelo ajuntamento de pares ou de nós, com base na distância entre eles, até todos estarem fundidos num número desejado de agrupamentos (Gutierrez-Osuna, 2002; Nikolaou et al., 2012).

No artigo de (Miljković, 2017), é possível verificar que com este algoritmo, é possível gerar conhecimento útil sobre o consumo de energia num edifício. No entanto, é mais provável conseguir produzir resultados úteis quando tem um número pequenos de agrupamentos (Gutierrez-Osuna, 2002).

### 2.5. MOTIF

Recentemente Su, McBride, & Goulias (2020) desenvolveram um novo método para identificar padrões baseado na mobilidade das pessoas e na sequência de atividades. Combina a análise baseada na rede *motif* e na análise baseada na sequência de atividades.

Neste artigo, para os autores testarem este novo método, decidiram encontrar padrões com base em deslocação, sendo elas: casa, trabalho, escola e compras, e no número de deslocação efetuadas. Foi usada uma base de dados da *2017 National Household Travel Survey* de um ano inteiro, entre 2016 e 2017, que continham também as características dos indivíduos, como sexo e idade. Inicialmente, com a rede *motifs*, foi efetuada uma divisão das pessoas em 5 grupos, no qual foi possível verificar que pessoas com características diferentes, pertencem em grupos *motifs* diferentes. De seguida, dentro de cada grupo, foi efetuada uma análise de sequência e agrupamento para identificar horários de atividades e viagens distintos. Para a segunda parte foi feita uma análise, em que combinaram a rede *motifs* com identificação dos padrões de alocação de tempo, em que usa análise de sequência e agrupamentos hierárquicos para mostrar distinções dentro de cada grupo *motifs*. Foi possível encontrar três padrões típicos de dias: dia em casa, dia de trabalho/dia de escola e outros sete padrões não típicos, mas com semelhanças entre os grupos *motifs* na sua pagada temporal (Su et al., 2020).

No entanto, este estudo não incluiu as pessoas que não se deslocaram pelo menos uma vez durante o dia, logo poderia estar a excluir o consumo das pessoas que trabalham em casa. Foi só baseada nas deslocações e não foi possível incluir os feriados. Um dos grupos *motifs* encontrados foram os alunos e trabalhadores a tempo integral que normalmente recaem num *motifs* de dois nós (Casa-Trabalho, Casa) ou (Casa-Escola, Casa). Por outro lado, pessoas que trabalham em tempo parcial ou pessoas mais jovens, tendem a visitar mais de dois locais por dia do que as pessoas mais velhas (Su et al., 2020).

## **2.6. AGRUPAMENTO K-MEANS**

O agrupamento *K-Means* é um outro método de agrupamento (*clustering*) e muito utilizada para analisar o consumo em edifícios. Este algoritmo, divide os dados de entrada em grupos não hierárquicos com similaridade, *clusters*, isto é, agrupados com base num critério específico. Este critério é quantificado pela distância euclidiana, que é uma distância entre dois pontos que pode ser provada pela aplicação sucessiva do teorema de Pitágoras. Apenas os melhores resultados depois das comparações feitas é que sairão como resultados deste algoritmo (Nikolaou et al., 2012; Wei et al., 2018)

Para este algoritmo funcionar, é necessário saber com antecedência o número do *cluster* *K* e as posições iniciais necessárias dos centros do *cluster*. É possível melhorar a viabilidade

deste algoritmo, usando os métodos modificados *Fuzzy*. Estes métodos permitem realizar agrupamentos suaves, ou seja, cada ponto de dados pode pertencer potencialmente a vários clusters e um grau de adesão é definido para caracterizar os relacionamentos. Os dois tipos de métodos *Fuzzy* são: *Fuzzy K-Means* e *Fuzzy C-Means* (FCM) (Wei et al., 2018).

## **2.7. REDE NEURONAL ARTIFICIAL**

A Rede Neuronal Artificial (ANNs) imita a arquitetura básica do cérebro humano e a sua unidade de processamento chama-se neurónio. É muito usado como método de previsão de consumo de energia em edifícios. Este algoritmo é capaz de aprender a relação entre os dados de entrada e capturar a informação essencial deles. Também tem tolerância a falhas, é robusto e imune a ruído. É capaz de resolver problemas não lineares. No processo de aprendizagem/treino, é capaz de aproveitar dados históricos, o que é essencial para o seu funcionamento e é necessário ser realizado antes de o aplicar em situações reais. No entanto, a arquitetura e a otimização da aprendizagem deste processo ainda está em desenvolvimento (Wei et al., 2018)

## **2.8. MODELO HÍBRIDO**

Fan & Chen (2006) propuseram um modelo híbrido baseado nos Mapas Auto-Organizados e na Máquina de Suporte de Vetores para preverem consumos a curto prazo. As vantagens apresentadas neste modelo é a capacidade de fazer face à não estacionariedade nas séries temporais de carga da eletricidade, o que tido cada vez mais efeitos no mercado. Pode tratar dias normais e semelhantes com diferentes esquemas. Também possui robustez e pode ser modificado para diferentes sistemas de energia ou mercados.

## **2.9. MODELOS ESCONDIDOS DE MARKOV**

Os Modelos Escondidos de *Markov* (HMMs) são úteis para achar probabilidades do estado estacionário de longo prazo no sistema, tendo em conta um número finito de estados possíveis, usando probabilidades de transição entre estados.

É necessário treinar este modelo. Ullah et al. (2018) utilizou dados do consumo de energia em edifícios durante um ano na Coreia do Sul, Seul, para o treino, validação e análise de resultados. Foi possível verificar que o HMM proposto pelos autores, obtiveram melhores resultados que outros algoritmos bem conhecidos, como o ANN e SVM. Também foi

possível confirmar que os HMMs são um dos melhores métodos para prever padrões do que outros algoritmos baseados em inteligência artificial.

Uma grande diferença entre os dados de consumo de energia comerciais e residências, é que os dados residenciais têm uma falta de granularidade e são geralmente recolhidos por uso de utensílios (Edwards et al., 2012).

Os sistemas de gestão de energia nos edifícios podem ser classificados em duas categorias:

- *Modelação Inversa*: Captura as dependências entre os diferentes tipos de dados de entrada e saída para criar uma fórmula matemática. Normalmente requer um grande conhecimento anterior;
- *Modelação Forward*: Tenta criar um horário ótimo de operação com requisitos de energia, ao usar parâmetros do design estrutural do edifício, meteorologia e informações ambientais (Ullah et al., 2018).

O uso de HMM é muito útil em dados temporais em série, mas não há muita investigação à sua aplicação em energia dos edifícios (Ullah et al., 2018).

Os autores Fang, Zhang, Liu e Shen (2018) utilizaram o HMM de forma a analisar os padrões de consumo de energia em edifícios residenciais. Os autores consideraram que com o avançar das tecnologias de comunicação, *smart grid* e *smart metering*, é possível extrair mais dados para fazer as análises. Neste caso, o HMM foi usado para aprender o comportamento da dinâmica de consumo em diferentes tipos de ambientes e dessa forma conseguirem o padrão de consumo individual de cada habitação. Neste artigo em específico, os autores realizam uma fusão dos dados de eletricidade e os dados de meteorologia, pois consideraram, pelas pesquisas efetuadas, que existe uma relação entre os padrões de consumo e os fatores de impacto (clima e características do edifício). Neste artigo, é assumido que o consumo de energia está relacionado com o aquecimento e arrefecimento da habitação. Por isso, foi definido três estados: aumento do consumo, diminuição do consumo e temperatura igual.

No artigo de Lin et al. (2018), foi utilizado o HMM para reconhecer os padrões de consumo para diferentes tipos de utilizações, sendo eles:

- Utensílios únicos para um único utilizador;
- Vários utensílios para um único utilizador;

- Vários utensílios para vários utilizadores.

Nos recentes estudos de consumo de energia em edifícios residenciais, o uso de casas inteligentes para este efeito foi uma opção escolhida por muitos. O artigo desenvolvido por Guenounou et al. (2021) considera que o consumo de energia está relacionado com os hábitos individuais. Neste artigo é usado um *machine learning* baseado em HMM, que permite monitorizar as atividades diárias relacionadas com o consumo de energia, utilizando, por exemplo, sensores pela habitação. Foi definido que os estados escondidos seriam as atividades dos utilizadores e os estados observáveis seriam os dados recolhidos pelos sensores. Os dados usados para este estudo estão presentes na literatura, em que vários estudos foram realizados com os mesmos conjuntos de dados. Os autores deste artigo conseguiram concluir que o HMM é um método eficaz, pois os resultados foram semelhantes aos outros estudos em que foram utilizados os mesmos dados.

Um outro artigo que usa habitações inteligentes para recolher dados é o artigo de Crandall e Cook (2010). Um problema que os autores quiseram ultrapassar com este artigo é a distinção dos consumos individuais entre as pessoas que vivem na mesma habitação. Uma forma que eles decidiram tentar fazer esta divisão foi ao usar o HMM. Para este estudo, foi utilizado duas bases de dados reais com o histórico dos habitantes. O objetivo principal foi conseguir implantar o HMM nas próprias habitações, por isso a sua compatibilidade com hardwares limitados, a memória e a eficácia são pontos muito importantes. Os autores representaram os diferentes residentes nos estados escondidos, enquanto os estados observáveis foram os sensores existentes na casa inteligente. Com o estudo realizado, foi possível concluir que o HMM teve um bom desempenho. No entanto requer mais sessões de treino para ter uma elevada precisão.

## 2.10. DISCUSSÃO

De acordo com a análise efetuada nos pontos anteriores, elaborou-se a tabela 2 para se poder identificar, resumidamente, as principais vantagens e desvantagens dos métodos/algoritmos analisados.

Tabela 2 - Vantagens e Desvantagens dos Métodos/Algoritmos

<b>Método</b>	<b>Vantagens</b>	<b>Desvantagens</b>
<b>SOM (Germano, 1999)</b>	Fácil de compreender; São bons a classificar dados.	Treino lento; Não é intuitivo; Problema às vezes de distinguir dados muito parecido.
<b>PCA (Kumar, 2019)</b>	Melhora a desempenho de algoritmos; Reduz o tamanho dos dados; Melhora a visualização dos dados.	Varáveis independentes ficam menos interpretáveis; É necessário padronizar os dados antes de usar o PCA; Perda de dados.
<b>MLPs (Gutierrez-Osuna, 2002)</b>	Aprender regressões não lineares complexas; Adição de camadas ocultas, aumenta a eficiência.	O número apropriado de unidades ocultas, ainda não é conhecido.
<b>Agrupamentos Hierárquicos (Bock, 2021; Gutierrez-Osuna, 2002)</b>	Fácil de entender.	Produz melhores resultados com um número menor de <i>clusters</i> ; Explicar os critérios de distância e de ligação; Falta de dados pode levar ao não funcionamento.

<b>Mofis (Su Et Al., 2020)</b>	Encontra padrões com base na deslocação;  Possível usar com várias características.	Em alguns casos, o método não contabiliza algumas circunstâncias.
<b>Agrupamento K-Means (Wei Et Al., 2018)</b>	Existem modificações que tornam este algoritmo mais viável, como os métodos <i>Fuzzy</i> .	É necessário definir algumas posições iniciais e o número do <i>cluster</i> K.
<b>ANNS (Wei Et Al., 2018)</b>	Aprender relações entre sinais;  Resolve problemas não lineares.	Otimização e tempo de aprendizagem ainda em otimização.
<b>Modelo Híbrido (Fan &amp; Chen, 2006)</b>	É robusto;  Pode ser modificado;  Junção do SOM e SVM.	Dados para treino do SOM devem de estar de acordo com as características sistema.
<b>HMMs (Ullah Et Al., 2018)</b>	Várias aplicações, como deteção de padrões.	Não podem mostrar dependências entre os estados ocultos.

Após a pesquisa das diferentes opções de análise, identificação e previsão de padrões, decidiu-se recorrer-se aos Modelos Escondidos de Markov para o presente trabalho. Com base no estado da arte, já há alguns artigos que implementaram HMMs para a previsão de padrões de consumo e conseguiram obter bons resultados, sendo portanto, uma escolha promissora para este trabalho.



# 3. MODELOS ESCONDIDOS DE MARKOV

Os HMMs contêm estados escondidos, estados observáveis, probabilidades iniciais, probabilidades de transição e probabilidades de emissão. Os estados escondidos são estados internos do sistema que não costumam ser visíveis para o utilizador. Por outro lado, os estados observáveis são diretamente visíveis para o utilizador e podem ser usados para fazer uma estimativa inteligente sobre o estado escondido. A nível das probabilidades iniciais, estas fornecem informações sobre a probabilidade de o sistema estar em qualquer estado escondido num determinado instante. As probabilidades de transição são as probabilidades de um estado escondido mudar para outro estado escondido ou manter-se no mesmo. Por fim, as probabilidades de emissão expõem a relação entre os estados escondidos e observáveis. O HMM é uma versão avançada no modelo básico de Markov que é especialmente formulado para quando os estados do sistema não estão diretamente visíveis, mas os resultados, dependendo dos estados escondidos, são visíveis (Ullah et al., 2018). O HMM é usado como uma ferramenta estatística para modelar sequências generativas caracterizadas por um conjunto de sequências observáveis. A estrutura do HMM é capaz de realizar processos estocásticos onde os estados escondido são governados por um processo de Markov e/ou as sequências observáveis têm uma dependência probabilística subjacente (Ramanathan, 2006).

Os HMMs são baseados na ampliação das cadeias de Markov. As cadeias de Markov são um modelo que permite obter variáveis aleatórias, probabilidade de sequência e estados, no qual qualquer um deles consegue assumir um valor dentro de um conjunto definido, que pode ser símbolos, números ou palavras. Um fator muito importante para as cadeias de Markov é o estado atual. A previsão dos futuros estados está dependente do estado atual, uma vez que este método faz uma suposição muito forte com base no presente, enquanto que os estados anteriores não têm impacto. Um exemplo usado por Jurafsky e Martin (2020) é o da meteorologia, que explica que as cadeias de Markov só conseguiriam prever o tempo dos dias seguintes, com base no dia atual, sem ter a possibilidade de olhar para os dias anteriores. Por isso, a suposição de Markov é que o que importa para prever o futuro não é o passado, mas sim o presente.

As cadeias de Markov são muito úteis quando se pretende prever uma sequência de eventos observáveis, mas às vezes é preciso ter em conta os eventos escondidos, que não se consegue ver diretamente. Nesse caso, os Modelos Escondidos de Markov são os mais vantajosos, uma vez que permite usar estados observáveis e estados escondidos, que podem ser fatores probabilísticos no modelo (Jurafsky & Martin, 2020).

Um HMM de primeira ordem tem 2 suposições para simplificação:

- A probabilidade de um estado ( $e_i$ ), apenas depende do estado anterior;

$$P(e_i|e_1 \dots e_{i-1}) = P(e_i|e_{i-1}) \quad (1)$$

- A probabilidade da saída de uma observação ( $o_i$ ) depende apenas do estado que produziu essa observação (Jurafsky & Martin, 2020).

$$P(o_i|q_i) \quad (2)$$

Os autores em Jurafsky e Martin (2020) utilizaram um exemplo realizado por (Eisner, 2002) que explica este modelo num exemplo simples. Há um climatologista num futuro longínquo que está a realizar um estudo sobre a história do aquecimento global, mas não informações sobre a temperatura no verão do ano  $x$  no local  $y$ , mas existe uma lista de quantos gelados foram vendidos durante esse ano. O principal objetivo é usar essa lista para conseguir prever a temperatura de cada dia. Para tal, simplificou-se o problema, assumindo-se que apenas há dois tipos de dias, um dia quente ( $Q$ ) e um dia frio ( $F$ ). Quando se vende mais gelados é

considerado um dia quente, e se se vendeu pouco, é considerado um dia frio. Foram definidos como estados escondidos os dias  $Q$  e  $F$ , e os estados observáveis a quantidade de gelados vendidos.

O HMM pode ser usado para o reconhecimento de fala. O estado observável para este caso pode ser as palavras que se vê, enquanto os estados escondidos podem ser as *tags* de classe gramatical (Jurafsky & Martin, 2020).

O HMM enfrenta três problemas principais:

- O problema de probabilidade (*likelihood*);
- O problema de decodificação;
- O problema de aprendizagem.

### 3.1. CONSTITUIÇÃO DO HMM

Como se pode observar na equação 3, o  $A$  corresponde às probabilidades de transição, o  $B$  às probabilidades de emissão e o  $\pi$  às probabilidades do estado inicial (Ramanathan, 2006).

$$\lambda = (A, B, \pi) . \tag{3}$$

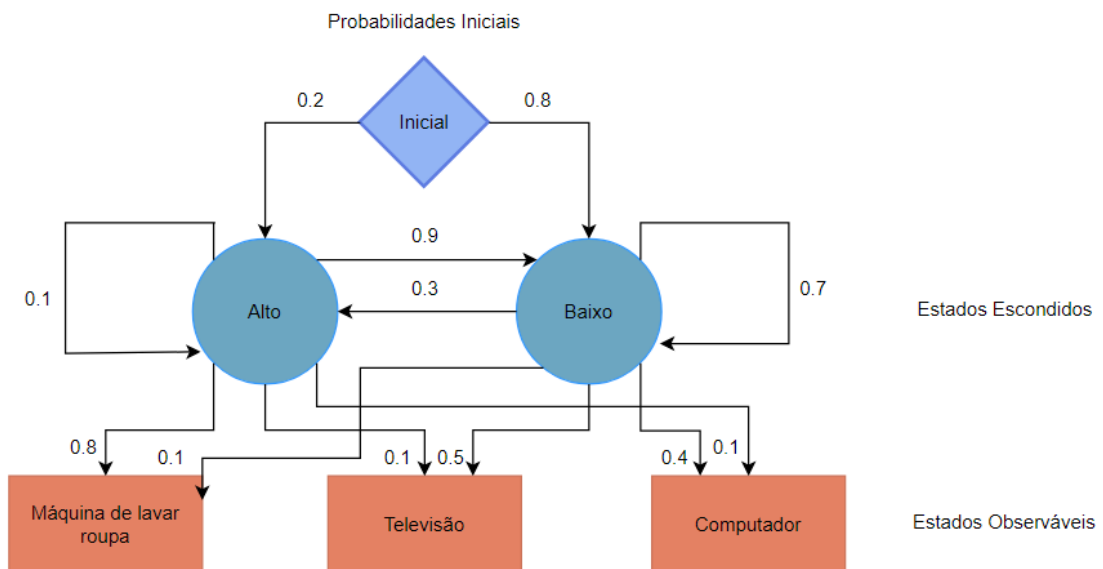


Figura 1 – Exemplo de HMM

O exemplo do HMM na figura 1 representa um edifício residencial que é parecido com o modelo das fases de teste, mas sofrendo algumas alterações dependendo do caso em concreto. Os estados escondidos,  $E = \{E1, E2\}$ , são o consumo de energia alto (estado  $E1$ ) e

baixo (estado  $E2$ ). Os estados observáveis,  $O = \{O1, O2, O3\}$ , são três utensílios presentes na habitação: máquina de lavar roupa (estado  $O1$ ), televisão (estado  $O2$ ) e computador (estado  $O3$ ).

As probabilidades iniciais são  $\pi = [0.2, 0.8]$ . As probabilidades de transição são  $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 0.1 & 0.9 \\ 0.3 & 0.7 \end{bmatrix}$ . As probabilidades de emissão são  $B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.5 & 0.4 \end{bmatrix}$ .

## 3.2. PROBLEMA DE PROBABILIDADE

Para resolver o problema de probabilidade, existem dois métodos principais que se pode utilizar: o algoritmo *Forward* e o algoritmo *Backward* (Burlando, 2018).

### 3.2.1. ALGORITMO FORWARD

O algoritmo *Forward* é constituído por três passos. O primeiro passo é a inicialização, no qual a primeira variável *forward* é calculada ao multiplicar a probabilidade inicial de um estado  $i$  pela probabilidade de emissão  $b$  de uma observação no instante 1, como mostra a equação 4 (Burlando, 2018).

$$\alpha_1(i) = \pi_i * b_i(O_i) \quad (4)$$

O segundo passo é a recursão. É usada a equação de recursão, equação 5, que nos fornece a variável seguinte do estado  $j$  como produto da variável *forward* anterior do estado  $i$ , multiplicada pela probabilidade de transição entre os estados  $i$  e  $j$  e a probabilidade de emissão  $b$  do estado  $j$  para a observação  $O$  (Burlando, 2018).

$$a_{t+1}(j) = \sum_{i=1}^N a_t(i) * a_{ij} * b_j(O_{t+1}) \quad (5)$$

O último passo é a terminação que é dada pela equação 6. Essa equação permite encontrar a probabilidade de uma sequência de observações  $O$  oriundo de um HMM  $\lambda$  onde é necessário somar todas as variáveis futuros do tempo  $T$  (Burlando, 2018).

$$P(O|\lambda) = \sum_{i=1}^N a_T(i) \quad (6)$$

### 3.2.2. ALGORITMO *BACKWARD*

Tanto como o algoritmo *forward*, o *backward* também é constituído pelos três passos, a inicialização, a recursão e a terminação. Este algoritmo não é usado para achar uma solução para o problema de probabilidade, mas é usado para indicar se o algoritmo *forward* funcionou corretamente e também poderá ser útil para resolver o problema de aprendizagem (Burlando, 2018).

A inicialização é adquirida com uma equação simples, equação 7, que significa que todas as variáveis *backward* de todos os estados são 1 no tempo  $T$  (Burlando, 2018).

$$B_T(i) = 1 \quad (7)$$

Para a recursão, a equação 8 indica somar todas as probabilidades de transição do estado  $i$  para o próximo estado  $j$ , multiplicando com a probabilidade de emissão da observação  $O$  no instante  $t+1$  do próximo estado e a variável *backward*  $\beta$  do próximo estado  $j$  em  $t+1$  (Burlando, 2018).

$$B_t(i) = \sum_{j=1}^N a_{ij} * b_j(O_{t+1}) * \beta_{t+1}(j) \quad (8)$$

A terminação é dada pela equação 9, no qual é o produto das somas das probabilidades iniciais  $\pi$  do estado  $i$ , com a probabilidade de emissão  $i$  pela observação no instante 1 do estado  $i$  e a variável *backward* no instante 1 do estado  $i$  (Burlando, 2018).

$$P(O|\lambda) = \sum_{i=1}^N \pi_i * b_i(O_1) * \beta_1(i) \quad (9)$$

### 3.3. DESCODIFICAÇÃO DO HMM

Como o HMM contém variáveis escondidas, para determinar a sequência de variáveis que é a causa de uma sequência de observações, é necessário realizar uma decodificação. A decodificação serve para determinar qual é a probabilidade da sequência de observações que é mais provável para um HMM  $\lambda = (A, B)$  e uma outra sequência de observações. Um dos algoritmos mais usados para a decodificação do HMM é o algoritmo *Viterbi* (Jurafsky & Martin, 2020).

O algoritmo *Viterbi* funciona como um programa dinâmico, que permite resolver problemas complicados ao dividi-los em problemas mais simples de uma forma recursiva. Assemelhasse ao algoritmo *forward* uma vez que tem os seus três passos: inicialização,

recursão e terminação. No entanto, tem um passo extra que é uma etapa de retrocesso (Burlando, 2019).

Para a inicialização  $\delta$ , é necessário multiplicar a probabilidade inicial do estado  $i$  com a probabilidade de emissão entre o estado  $i$  para a observação no instante 1, como se pode ver pela equação 10. Para se obter a sequência de estados, é necessário seguir o argumento que se maximizou para cada  $t$  e  $j$ , que é feito com o vetor  $\psi$ , que é inicializado em 0, como se pode ver na equação 11 (Burlando, 2019).

$$\delta_1(i) = \pi_i * b_i(O_1) \quad (10)$$

$$P\psi_1(i) = 0 \quad (11)$$

Para a recursão, é necessário efetuar multiplicações e de seguida encontrar qual foi o valor máximo obtido. Como se pode ver pela equação 12, é preciso multiplicar a variável *viterbi* anterior do estado  $i$ , com a probabilidade de transição entre os estados  $i$  e  $j$ , e a probabilidade de emissão entre o estado  $j$  e a observação. Para o vetor  $\psi$ , é adicionado o argumento dado pela equação 13 (Burlando, 2019).

$$\delta_t(j) = \max_{i=1}^N \delta_{t-1}(i) * a_{ij} * b_j(O_t) \quad (12)$$

$$\psi_t(j) = \arg \max_{i=1}^N \delta_{t-1}(i) * a_{ij} * b_j(O_t) \quad (13)$$

A terminação consiste na equação 14 que mostra, com base nos parâmetros do HMM e observações, a probabilidade de toda a sequência de estados até ao instante  $T+1$  ter sido produzida. Para isso, é necessário encontrar o valor máximo obtido de todas as variáveis *viterbi* calculadas nos instantes  $T$ . O último elemento do vetor  $\psi$  é dado pela equação 15 (Burlando, 2019).

$$\delta_{T+1}(j) = \max_{i=1}^N \delta_T(i) \quad (14)$$

$$\psi_{T+1}(j) = \arg \max_{i=1}^N \delta_T(i) \quad (15)$$

O último passo da decodificação ao usar o algoritmo *viterbi* é o retrocesso, no qual  $\psi$  fornece o último estado da sequência de estados escondidos obtidos que vai ser usado para começar esta etapa. A equação 16 mostra como se encontra a sequência de estados escondidos, seguindo do fim para o início o vetor  $\psi$  (Burlando, 2019).

$$q_t^* = \psi_{t+1}(q_{t+1}^*) \quad (16)$$

### 3.4. TREINO DO HMM

Um modelo usado para treinar o HMM é chamado de *Viterbi Training*. O *Viterbi* usado para descodificação é diferente do *Viterbi* usado para o treino. Este modelo procura maximizar a probabilidade da sequência dos estados escondidos mais prováveis. O *Viterbi Training*, consegue recolher a sequência dos estados escondidos que são dados à sequência de estados observáveis, quando é aplicado computacionalmente. Também é capaz de estimar os parâmetros dos modelos. O *Viterbi Training* tem um problema de consistência quando comparado com o modelo *Maximum Likelihood Estimation*, mas tem uma boa performance em aplicações práticas, o que explica a sua alta utilização (Allahverdyan & Galstyan, 2011).

Um outro treino que é possível aplicar ao HMM é o algoritmo *Baum-Welch*. Este algoritmo é usado quando não se sabe as probabilidades de transição e emissão do modelo. Neste caso, o algoritmo fornece informações possíveis dessas probabilidades.

### 3.5. MODELO COMPUTACIONAL

Para aplicar o HMM num problema prático, é necessário recorrer a um programa que contenha os elementos básicos do HMM, que são os estados escondidos, estados observáveis, probabilidades iniciais, probabilidades de transição e de emissão. Além, disso, também é importante que o programa consiga realizar treinos e conseguir prever dados.

Tendo isso em mente, foi utilizado a linguagem de programação R. Para a implementação do HMM, foi usado a livreria “HMM” Himmelman (2015), uma vez que permite fazer treinos e previsões de dados usando o Modelo Escondido de *Markov*.

Uma das principais funções da livreria “HMM” é o *initHMM*, que serve como um iniciador do HMM. Para esta função, será necessário definir os estados escondidos, estados observáveis, probabilidades iniciais dos estados escondidos, probabilidades de transição entre estados escondidos e probabilidades de emissão de estado escondido para estado observável. Relativamente ao treino, esta livreria oferece o treino *Viterbi*, *viterbiTraining*, que pode ser usado ao ter o HMM inicializado, um conjunto de observações e um número máximo de iterações, dando no final um modelo de HMM resultante do treino. Uma outra função importante é o *simHMM*, que pode ser utilizado para simular os estados escondidos

e observáveis de um HMM, no que necessita de um HMM inicializado e do tamanho de dados que se pretende simular, resultando nos estados escondidos e observáveis previstos.

## 4. CASO DE ESTUDO

Neste capítulo são descritos os dados levantados e usados para o desenvolvimento de um caso de estudo para o teste do HMM. É mostrado como foi feita a análise dos dados, os resultados obtidos e são tiradas as respectivas conclusões.

O principal objetivo dos casos de estudo é conseguir analisar um conjunto de dados de consumo de uma habitação residencial, de forma a conseguir prever corretamente dados futuros. No processo, é necessário obter as probabilidades necessárias para o funcionamento do HMM e definir os estados escondidos e observáveis para o problema. De seguida, treinar o HMM com um conjunto de dados históricos de forma a conseguir prever corretamente os dados futuros.

Durante o teste do programa, foi preciso efetuar diversos testes, de forma a conseguir refinar o modelo e o processo de treino, com vista à obtenção de resultados satisfatórios.

### **4.1. LEVANTAMENTO E TRATAMENTO DOS DADOS**

Para a realização deste trabalho, é muito importante ter uma boa base de dados para realizar o teste do HMM para a previsão e identificação de padrões de consumo.

Uma vez que o HMM requer algum treino para poder funcionar o melhor possível, decidiu-se procurar uma base de dados com muitas medições, para se poder ter um conjunto de treino

grande. Além disso, também é essencial que os dados tenham vários equipamentos com medições.

Tendo isso em mente, foi escolhida a base de dados *RETIF: Electrical Load Measurements* (Murray et al., 2017). Esta base de dados contém os dados de medição de 21 habitações residenciais. Para este caso de estudo, foi selecionado a habitação 1.

A habitação número 1 foi construída entre 1975 e 1980, tem 35 equipamentos elétricos e 4 quartos. Na base de dados, apenas 9 equipamentos mostraram o seu consumo individual em cada medição, mas o conjunto agregado de consumo, inclui os 35 equipamentos. Os 9 equipamentos com medições são os seguintes:

1. Equipamento 1 – Frigorífico – Marca: *Hotpoint* – Modelo: RLA50P;
2. Equipamento 2 – Congelador (1) – Marca: *Beko* – Modelo: CF393APW;
3. Equipamento 3 – Congelador (2) – Marca: Desconhecido – Modelo: Desconhecido;
4. Equipamento 4 – Máquina de Secar Roupa – Marca: *Creda* – Modelo: T522VW;
5. Equipamento 5 – Máquina de Lavar Roupa – Marca: *Beko* – Modelo: WMC6140;
6. Equipamento 6 – Máquina de Lavar Loiça – Marca: *Bosch* – Modelo: Desconhecido;
7. Equipamento 7 – Computador – Marca: *Lenovo* – Modelo: H520s;
8. Equipamento 8 – Televisão – Marca: *Toshiba* – Modelo: 32BL502b;
9. Equipamento 9 – Aquecimento Elétrico – Marca: *GLEN* – Modelo: 2172.

Os dados da habitação 1 não foram medidos de uma forma regular, mas têm em média 8 dados por minuto. No entanto, por motivos de limitações computacionais, foi necessário realizar uma redução ao realizar médias, para assim diminuir a dimensão dos dados.

Na figura 2, está presente a evolução temporal dos equipamentos com o consumo individual durante o dia 10 de outubro de 2013.

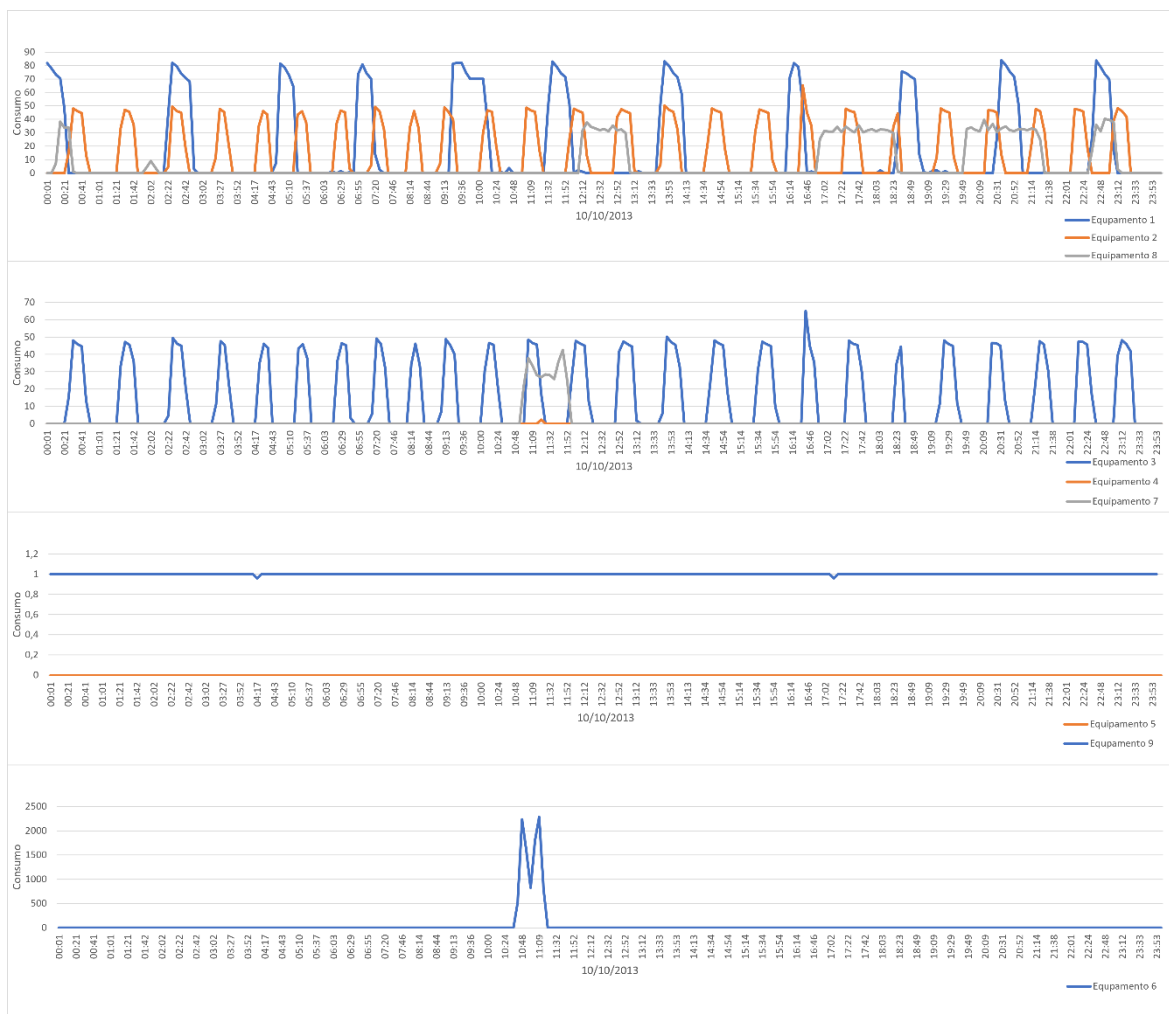


Figura 2 - Evolução temporal dos equipamentos num dia

#### 4.2. TESTE 1 – 4 ESTADOS ESCONDIDOS E 9 ESTADOS OBSERVÁVEIS

Para o uso do HMM, é necessário a definição dos estados escondidos, estados observáveis/observações, as probabilidades de transição entre estados escondidos e as probabilidades de emissão entre os estados escondidos e as observações.

Para o primeiro teste, decidiu-se criar 4 estados escondidos baseados no consumo agregado de energia, sendo eles:

- Estado 1 – Consumo entre 0 W e 236,7 W;
- Estado 2 – Consumo entre 236,8 W e 452,7 W;
- Estado 3 – Consumo entre 452,8 W e 1078,7 W;
- Estado 4 – Consumo maior que 1078,8 W.

É possível ver na figura 3 as percentagens de observações que pertencem a cada estado escondido.

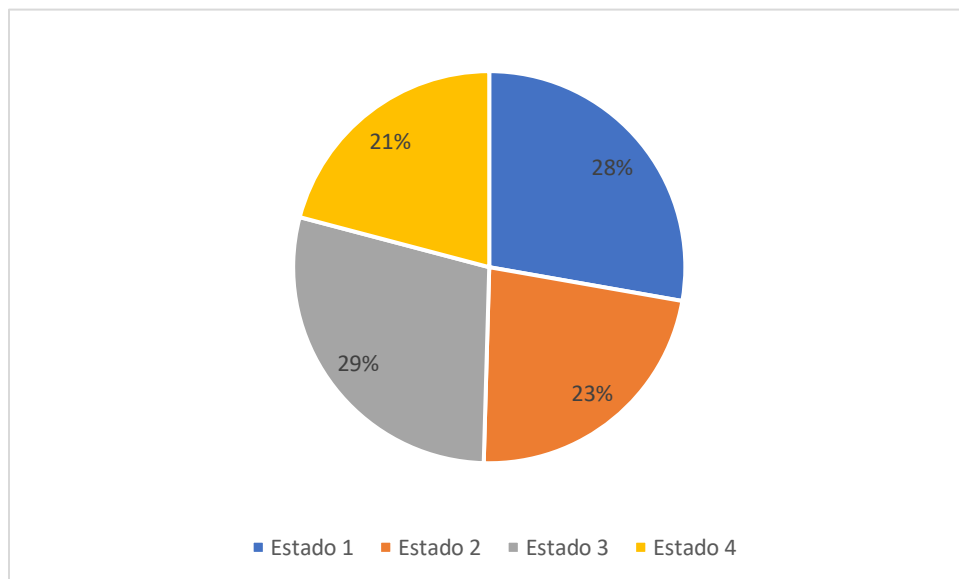


Figura 3 - Percentagem de cada estado escondido para o teste 1

Para os estados observáveis, decidiu-se que seriam os equipamentos elétricos, logo tem-se 9 estados observáveis, que correspondem aos equipamentos do 1 ao 9.

Para se poder iniciar o HMM, é necessário calcular as probabilidades de transição entre estados escondidos. Para isso, foi atribuído a cada medição um estado e de seguida foi-se verificar quantas vezes um estado mudava para outro estado, como por exemplo, do estado 4 para o estado 2. Tendo o número total de vezes que as mudanças foram efetuadas, dividiu-se pelo número total de medições, dando assim a probabilidade de transição entre cada estado,  $A$ , como mostra a figura 4.

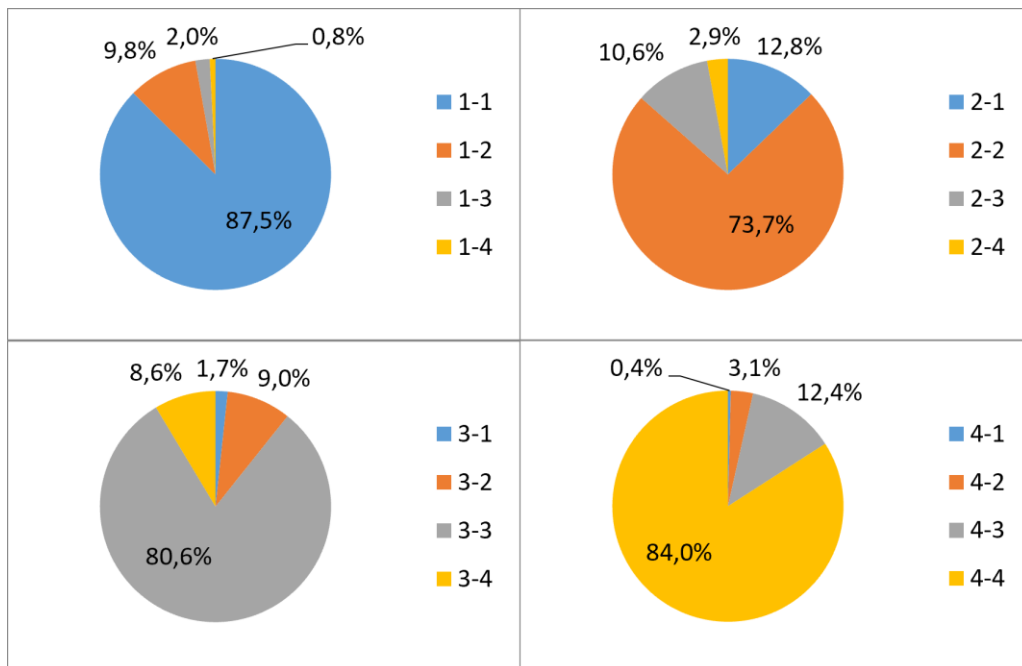


Figura 4 - Probabilidades de transição para o teste 1 (estado escondido - estado escondido)

Os estados observáveis neste teste são os equipamentos, tendo o HMM nove estados observáveis possíveis. Uma medição de um certo estado escondido, tem como estado observável o equipamento que consumiu mais energia.

Para calcular as probabilidades de emissão, decidiu-se verificar qual foi o equipamento elétrico que teve maior consumo naquela medição, por exemplo, uma medição que pertence ao estado escondido 3, teve o equipamento 7 com maior consumo, logo o 3-7 (estado-observação) tem um acontecimento. Depois de seguir o mesmo raciocínio para todas as medições, divide-se esse número pelo número total de medições, que são 20970, dando assim a probabilidade de emissão,  $B$ , entre o estado escondido e o estado observável, como mostra a figura 5.

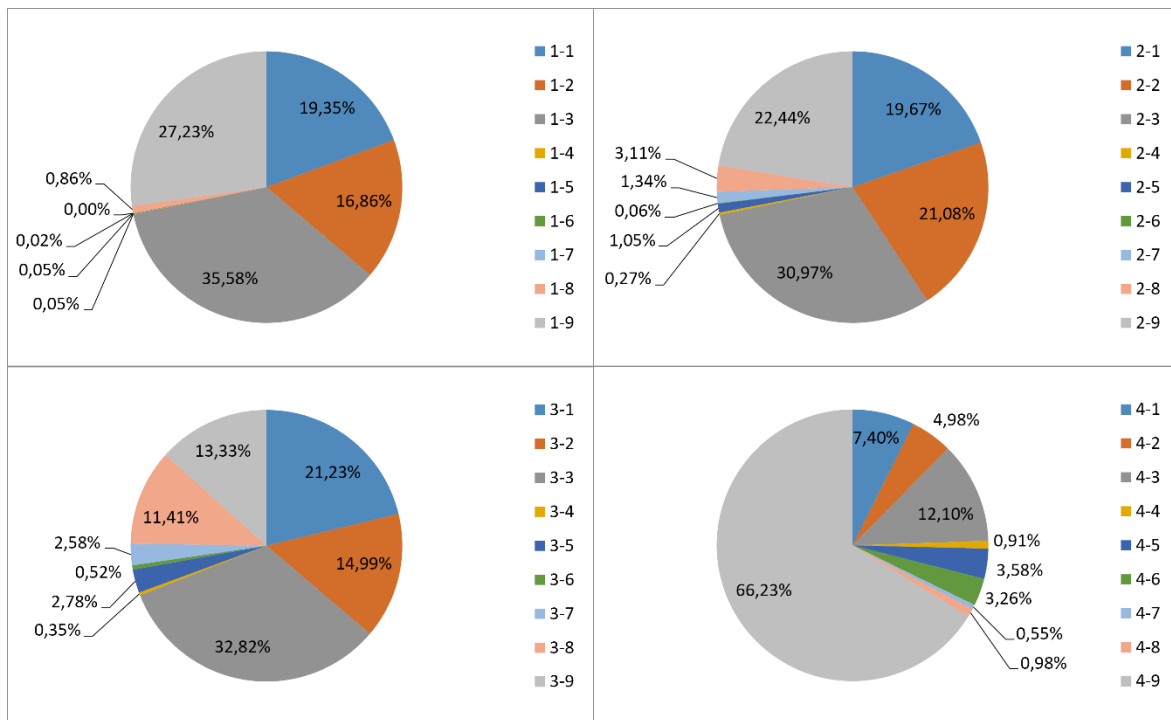


Figura 5 - Probabilidades de emissão para o teste 1 (estado escondido – estado observável)

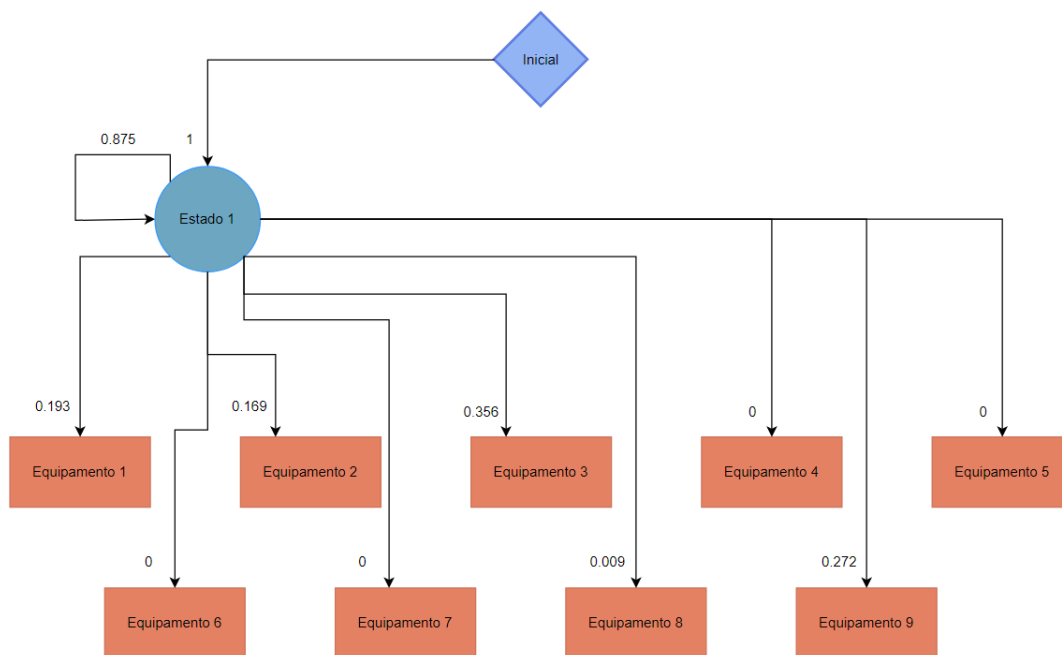


Figura 6 - Estado 1 do teste 1

Como o presente HMM tem muitos estados escondidos e estados observáveis, uma representação completa tornaria impossível a sua visualização, por isso, é possível ver na figura 6, uma parte da representação do HMM que está aplicada no estado 1. Está presente

a probabilidade de transmissão se se manter no mesmo estado escondido 1, e todas as probabilidades de emissão do estado escondido 1 para os nove estados observáveis.

Com o HMM iniciado, é necessário efetuar o seu treino. Foi escolhido o algoritmo *Viterbi* para realizar o treino, uma vez que se tem as probabilidades de transição e de emissão já calculadas. O treino *Viterbi*, quando comparado com o *Baum-Welch*, consegue fornecer a sequência completa mais provável com base no modelo, além de ser mais leve computacionalmente e mais rápida.

Normalmente, o treino é efetuado com 70% dos dados da base de dados e os restantes 30% são usados para testar o Modelo Escondido de Markov, mas para começar, realizou-se o teste para o último dia, usando todos os dados anteriores como treino.

A livreria utilizada no R, para o treino *Viterbi*, requer um HMM já inicializado, um conjunto de acontecimentos observáveis e o número desejado de iterações. Com o HMM inicializado, usou-se todos os estados observáveis anteriores com exceção do último dia. O número de iterações correspondeu ao número de medições existentes no último dia, que são 920.

Com os resultados tão pouco satisfatórios do *Viterbi*, como se pode verificar na figura 7.

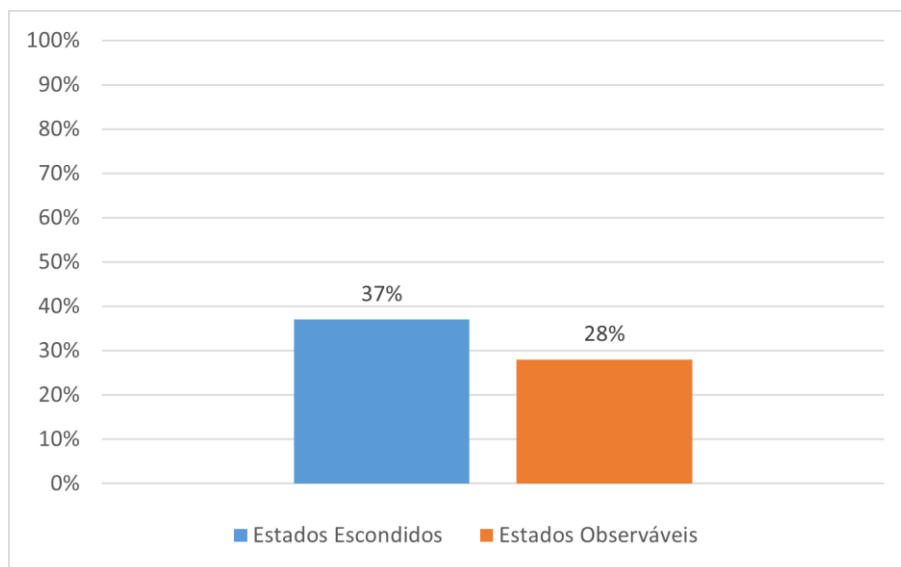


Figura 7 - Resultados do teste 1 com *Viterbi*

O HMM conseguiu prever corretamente 37% dos estados escondidos e 28% dos estados observáveis. Esta percentagem é o dividendo entre número de estados em que acertou e o número total de estados previstos, multiplicando depois por 100.

É possível ver na figura 8, a comparação dos estados escondidos e estados observáveis entre os estados reais e os estados previstos.

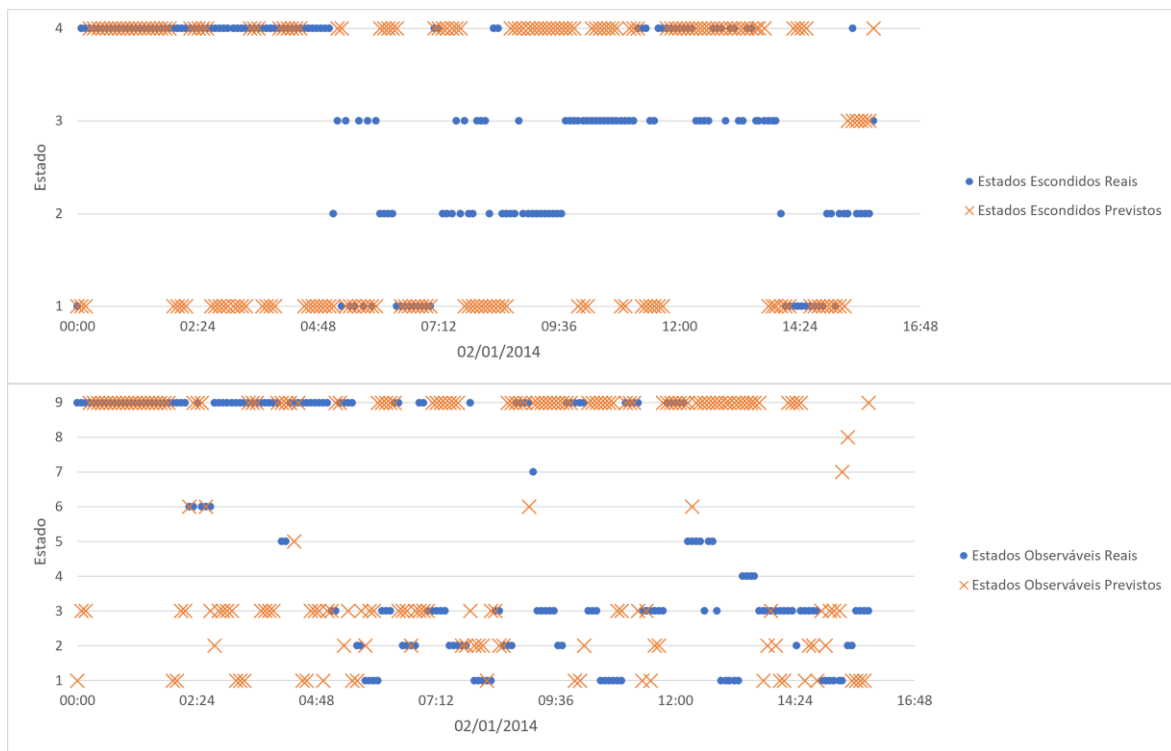


Figura 8 - Comparação dos estados escondidos e observáveis no teste 1

Utilizou-se o treino *Baum-Welch* para fazer uma comparação entre os dois algoritmos.

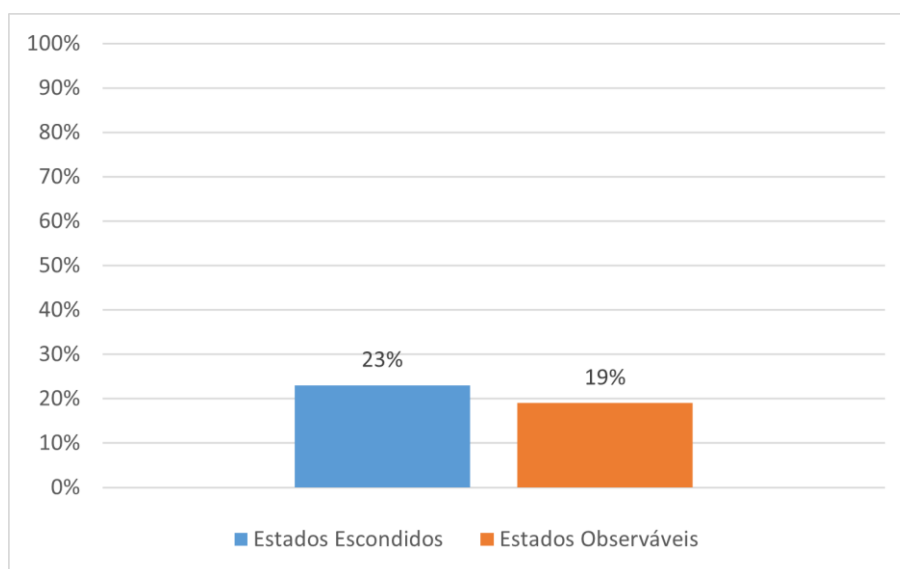


Figura 9 - Resultados do teste 1 com *Baum-Welch*

Os resultados apresentados na figura 9, mostram que o treino *Baum-Welch* foi ainda pior que o *Viterbi*, por isso, comprovou que o algoritmo mais apropriado é mesmo o treino *Viterbi*.

Apesar de o teste para o último dia não ter tido bons resultados, fez-se o recomendado 70% de treino e 30% de teste, uma vez que já se tinha o HMM inicializado e seria bom verificar os resultados.

Para o treino, utilizou-se as observações dos primeiros 70% das medições e realizou-se 6291 iterações, que correspondem as últimos 30%, como mostra a figura 10.

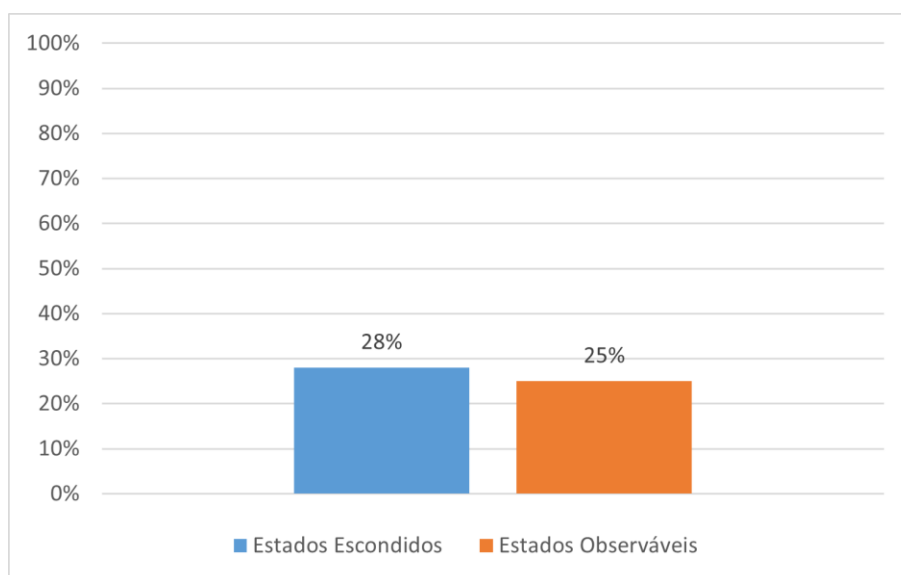


Figura 10 - Resultados do teste 1 para 70/30 com *Viterbi*

Os resultados do treino 70/30 foram igualmente maus, tendo conseguido prever corretamente 28% dos estados escondidos.

#### **4.3. TESTE 2 – 4 ESTADOS ESCONDIDOS A 25% E 9 ESTADOS OBSERVÁVEIS**

De forma a tentar melhorar os resultados obtidos, realizou-se um balanceamento dos dados relativo à distribuição do número de observações por estado escondido. Para isso procedeu-se a uma divisão dos dados, de forma a que o número de observações fosse igual para cada estado escondido, isto é, cada estado escondido ter aproximadamente 25% do total de medições, como mostra a figura 11. Para isso acontecer, cada estado precisou de ter os seguintes consumos:

- Estado 1 – Consumo entre 0 W e 230,7 W;
- Estado 2 – Consumo entre 230,8 W e 452,7 W;
- Estado 3 – Consumo entre 452,8 W e 910,7 W;
- Estado 4 – Consumo maior que 910,8 W.

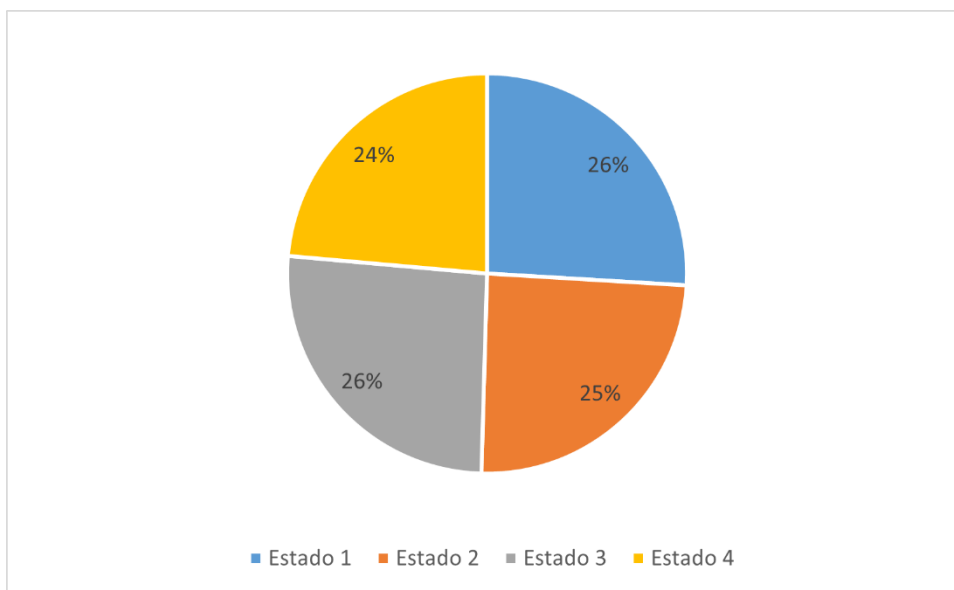


Figura 11 - Número de cada estado do teste 2

Para calcular as probabilidades de transição entre estados escondidos e as probabilidades de emissão do estado escondido para o estado observável, seguiu-se o mesmo raciocínio do teste 1. Pode-se verificar as probabilidades transição e emissão nas figuras 12 e 13, respectivamente.

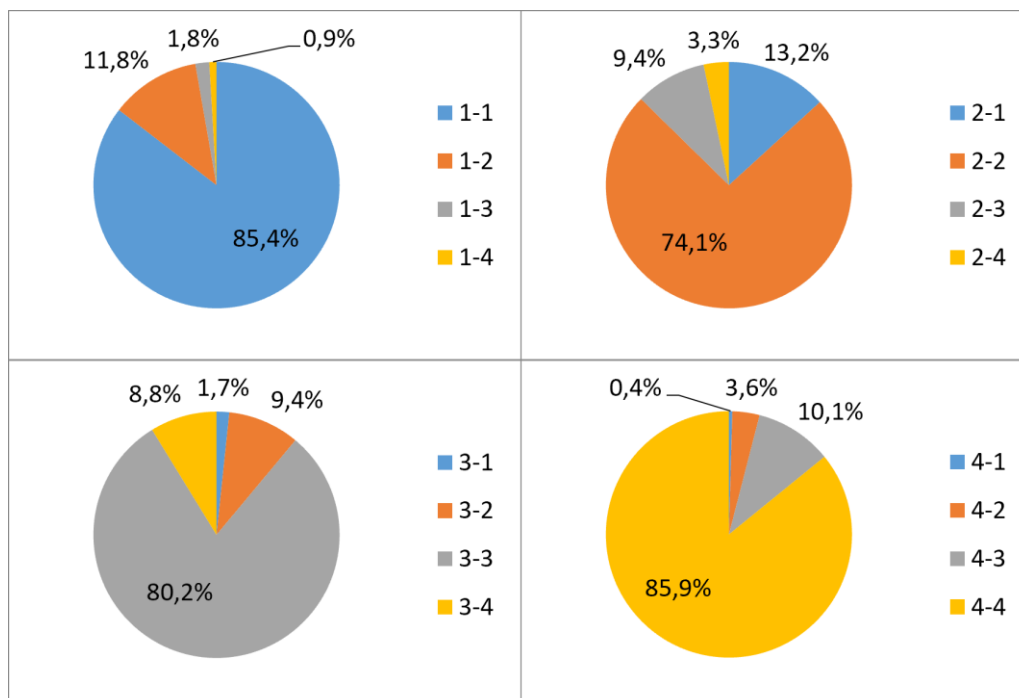


Figura 12 - Probabilidade de transição do teste 2

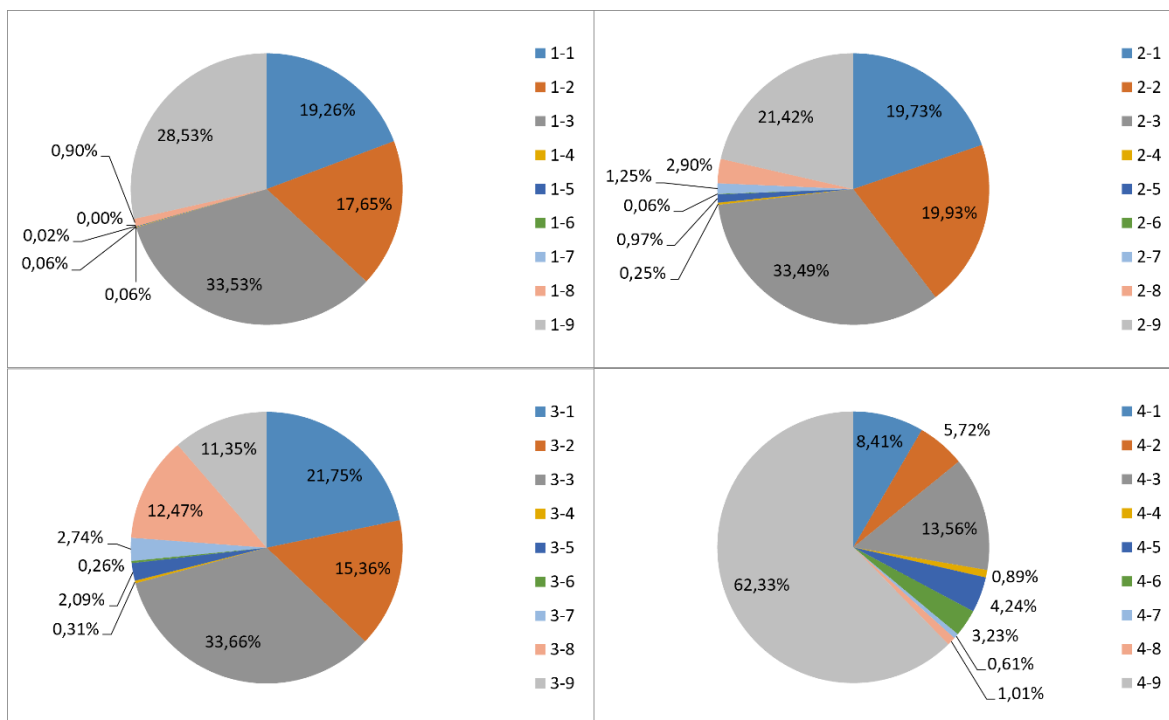


Figura 13 - Probabilidade de emissão para teste 2

Para o treino do teste 2, já com as probabilidades necessárias calculadas, seguiu-se para o treino de 70% e o teste de 30%.

Como se pode observar pela figura 14, os resultados neste teste foram melhores que os do teste 1.

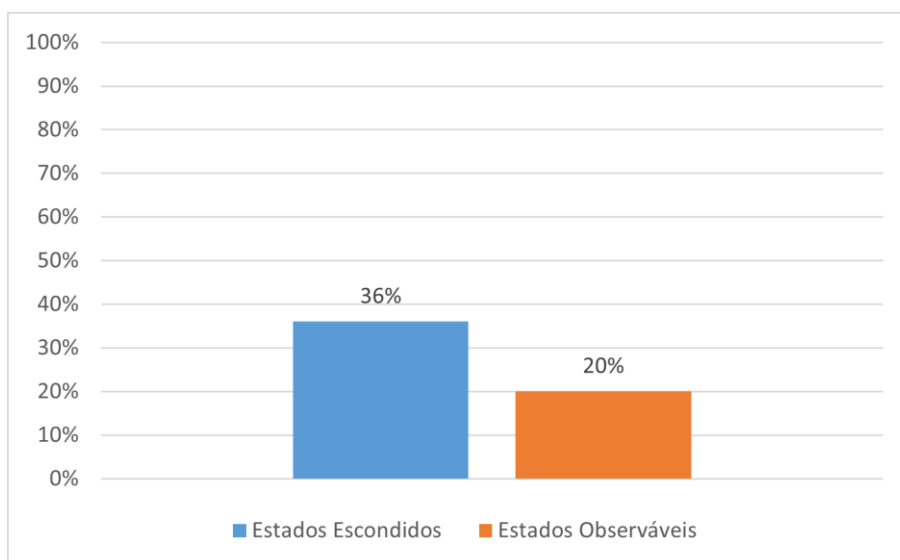


Figura 14 - Resultados do teste 2

Um dos principais fatores que pode ter levado para a melhoria dos resultados é o facto de os estados escondidos estarem divididos de uma forma quase igualitária, o que permite o HMM prever um padrão.

#### 4.4. TESTE 3 – 2 ESTADOS ESCONDIDOS E 2 ESTADOS OBSERVÁVEIS

Para o terceiro teste, resolveu-se tentar um problema mais pequeno. Uma possível causa para os maus resultados do HMM é a livreria do HMM no R não estar preparada para a grande quantidade de dados que se estava a utilizar anteriormente. Por isso, para o teste 3, criou-se apenas 2 estados escondidos, 2 estados observáveis e uma menor quantidade de medições, tendo 531 em vez das 20970 dos testes anteriores.

Para escolher os estados escondidos, decidiu-se simplificar muito o problema, sendo o estado 1 as horas entre as 07:00 e as 18:59 e o estado 2 as horas entre 19:00 e as 06:59. Para o estado observável, simplesmente foi se verificar se o consumo agregado era inferior a 250 W para o estado 1, ou igual ou superior a 250 W para o estado 2. As suas percentagens podem ser observadas na figura 15.

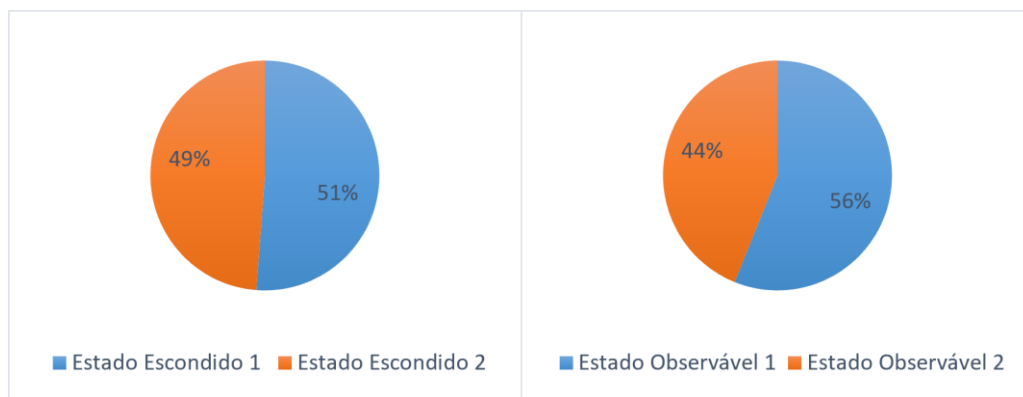


Figura 15 - Número dos estados escondidos e observáveis do teste 3

Para calcular as probabilidades de transição e emissão, seguiu-se o mesmo raciocínio que nos testes anteriores, estando presentes nas figuras 16 e 17, respetivamente.

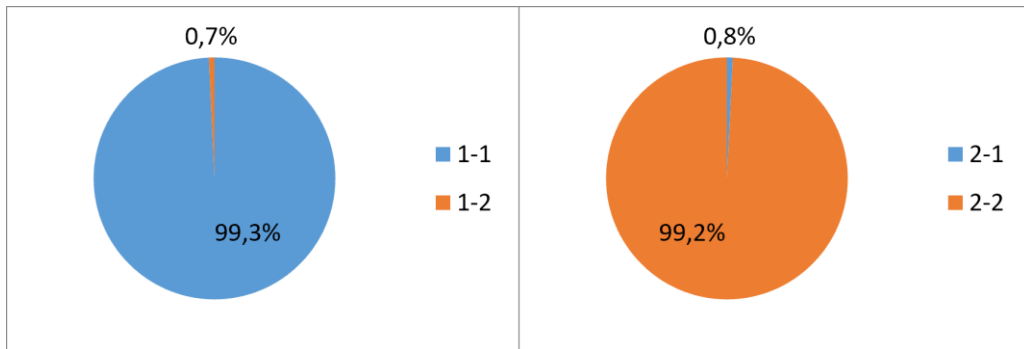


Figura 16 - Probabilidade de transição para teste 3

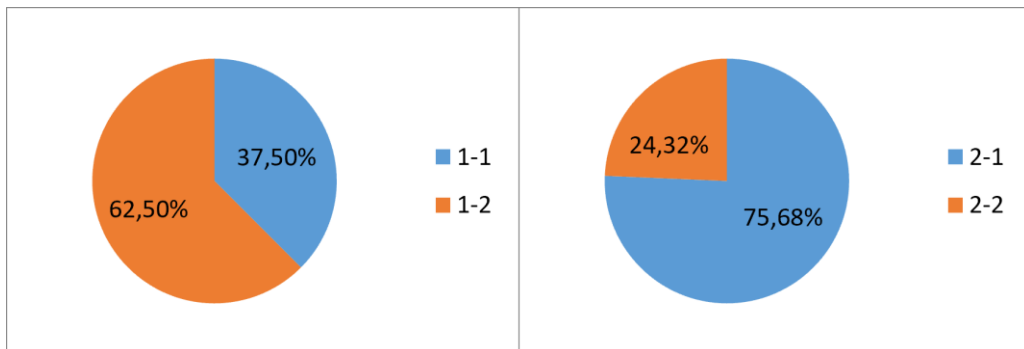


Figura 17 - Probabilidade de emissão para teste 3

Para o teste do HMM, fez-se o treino utilizando medições de 2 dias completos e para avaliar os resultados do HMM testou-se com os dados do dia seguinte.

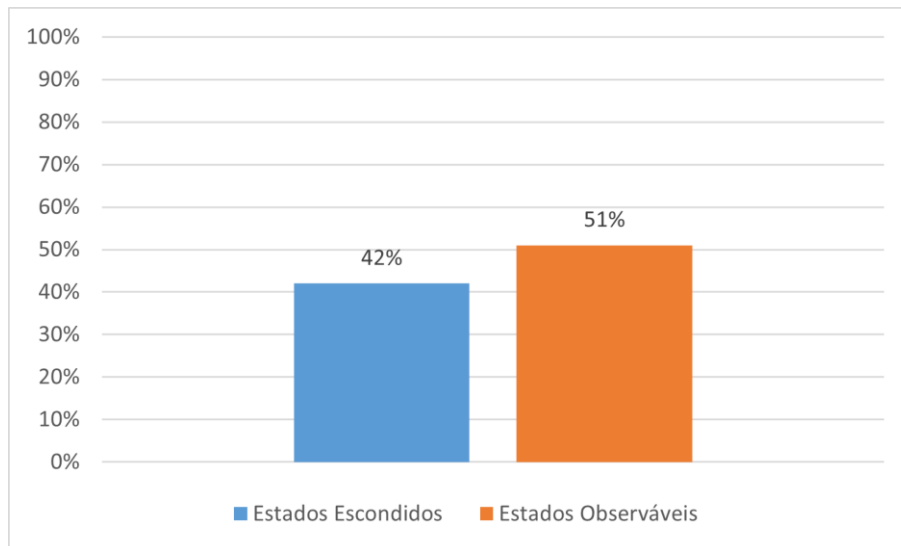


Figura 18 - Resultados do teste 3

Como se pode observar na figura 18, os resultados melhoraram imenso, quando comparado com os testes anteriores. Apesar de os resultados não serem muito satisfatórios, mostra que

uma possível causa para os resultados maus dos testes anteriores, é a grande quantidade de dados que o programa não é capaz de suportar.

Na figura 19, é possível ver a comparação entre os estados reais e previstos dos estados escondidos e observáveis.

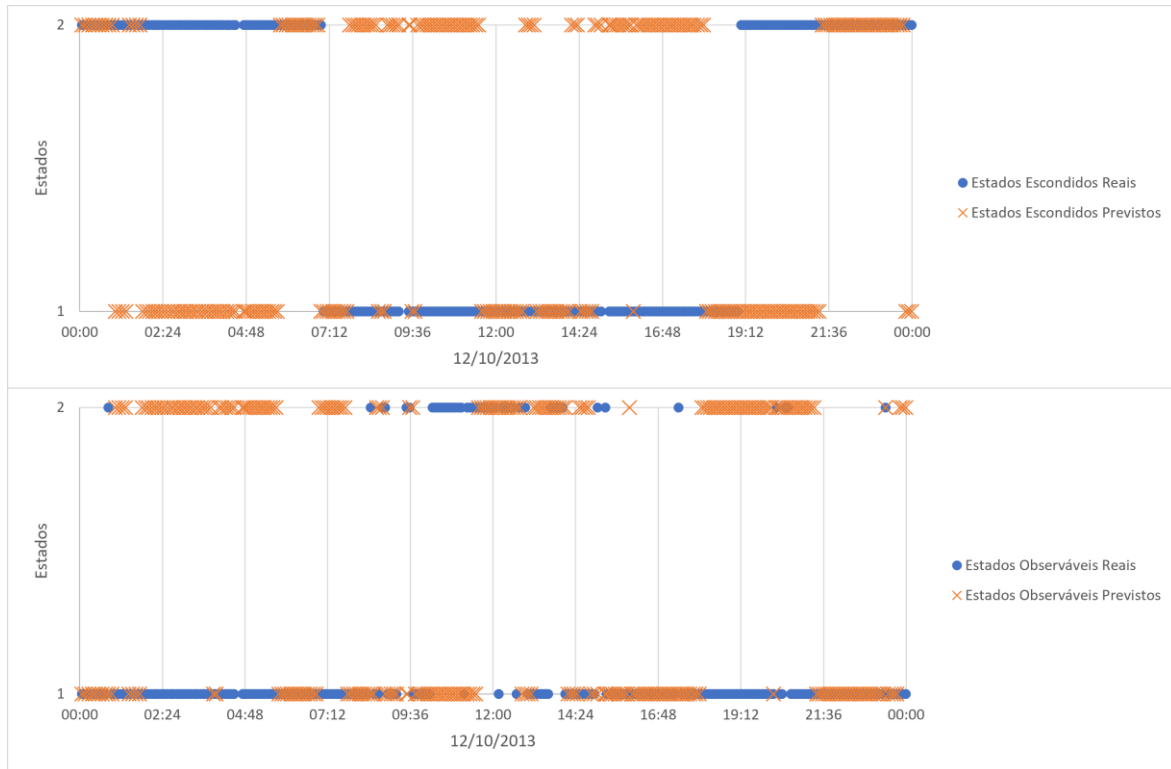


Figura 19 - Comparação dos estados escondidos e observáveis no teste 3

#### 4.5. TESTE 4 – 2 ESTADOS ESCONDIDOS E 2 ESTADOS OBSERVÁVEIS COM MENOS DADOS

Como os resultados do teste 3 mostraram uma melhoria significativa com a simplificação dos dados, decidiu-se diminuir ainda mais os dados, mas desta vez aumentar a diversidade dos mesmos. Em vez de ter os dados a cada 5/6 minutos, realizou-se médias para ter uma medição a cada 4 horas. Dessa forma, é possível calcular as probabilidades necessárias para iniciar o HMM com dados equivalentes a 7 dias, sendo no total 48 medições. O estado escondido 1 equivale às horas 00:00, 04:00 e 20:00, enquanto o estado escondido 2 equivale às horas 08:00, 12:00 e 16:00, representado uma separação entre dia e noite. Para os estados observáveis, seguiu-se a mesma regra que no teste 3. A figura 20 mostra a distribuição dos estados escondidos e observáveis.

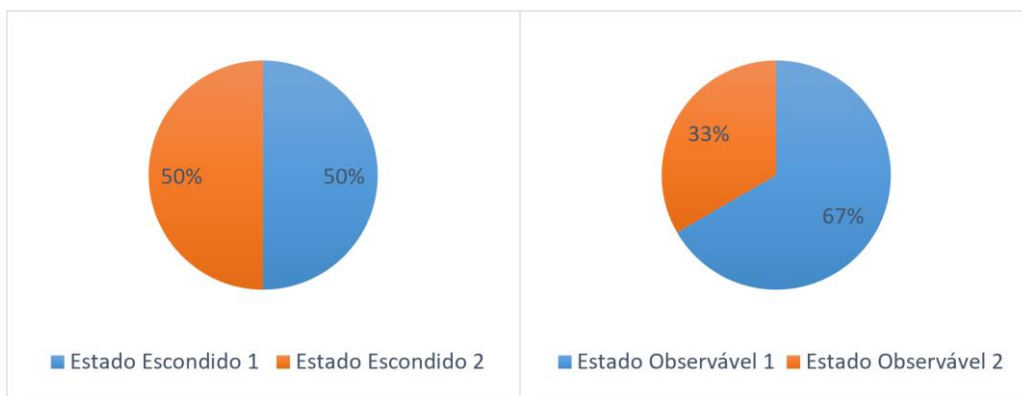


Figura 20 - Número dos estados escondidos e observáveis do teste 4

O modo do cálculo das probabilidades de transição e emissão são iguais ao do teste anterior e podem ser vistas na figura 21 e 22.

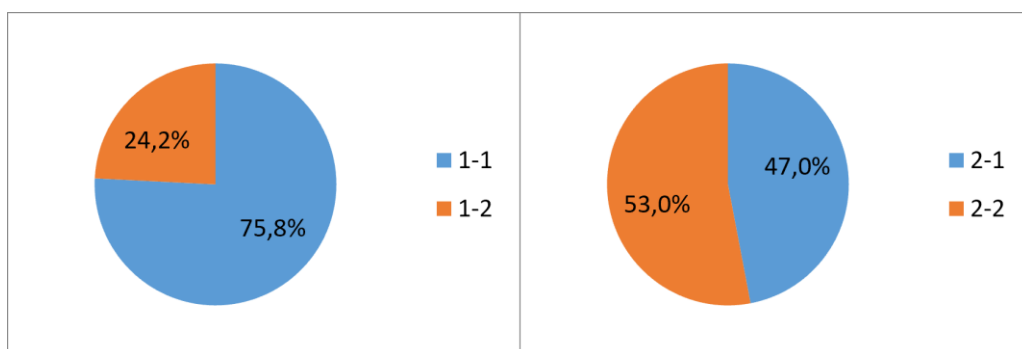


Figura 21 - Probabilidade de transição para teste 4

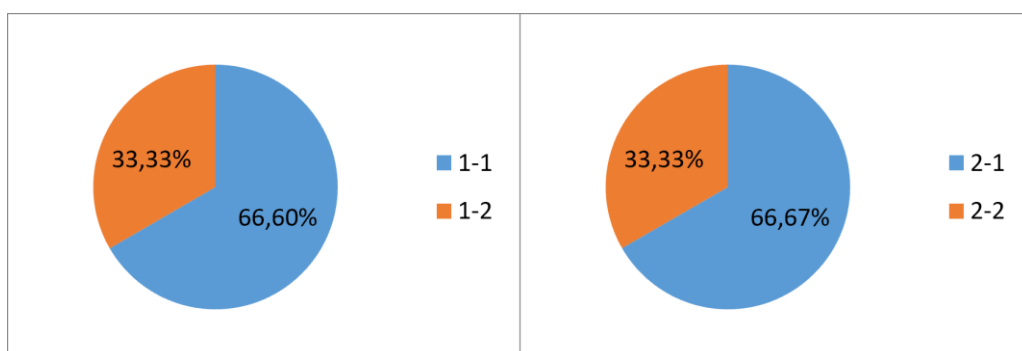


Figura 22 - Probabilidade de emissão para teste 4

Para realizar o treino do HMM, utilizou-se todos os dados usados para calcular as probabilidades de transição e emissão do HMM. Como teste utilizou-se uma previsão para os próximos dois dias.

Como mostra a figura 23, os resultados mostraram uma ligeira melhoria.

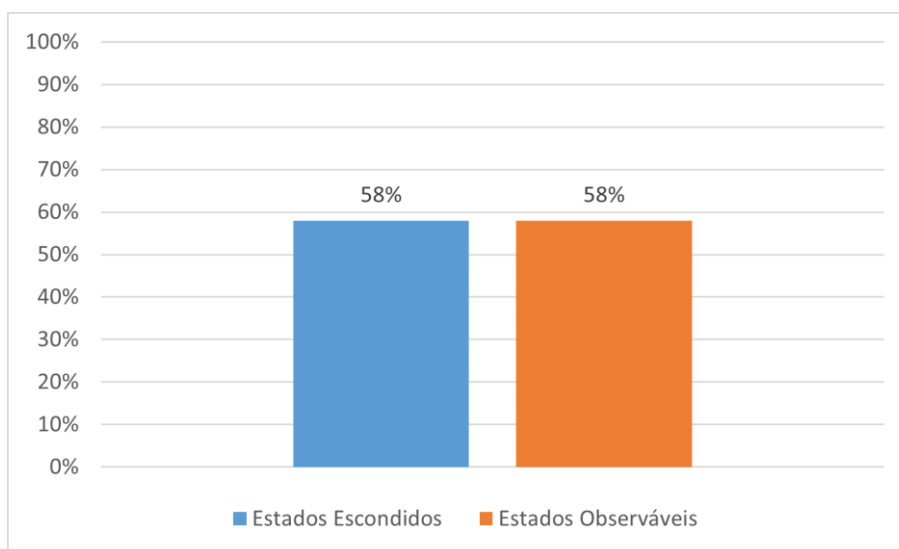


Figura 23 - Resultados do teste 4

Uma possível causa para tal é o facto de haver uma maior variedade de dias, que cria um melhor treino para o HMM.

Na figura 24, encontra-se a comparação entre os estados escondidos e estados observáveis reais e previstos no teste 4.

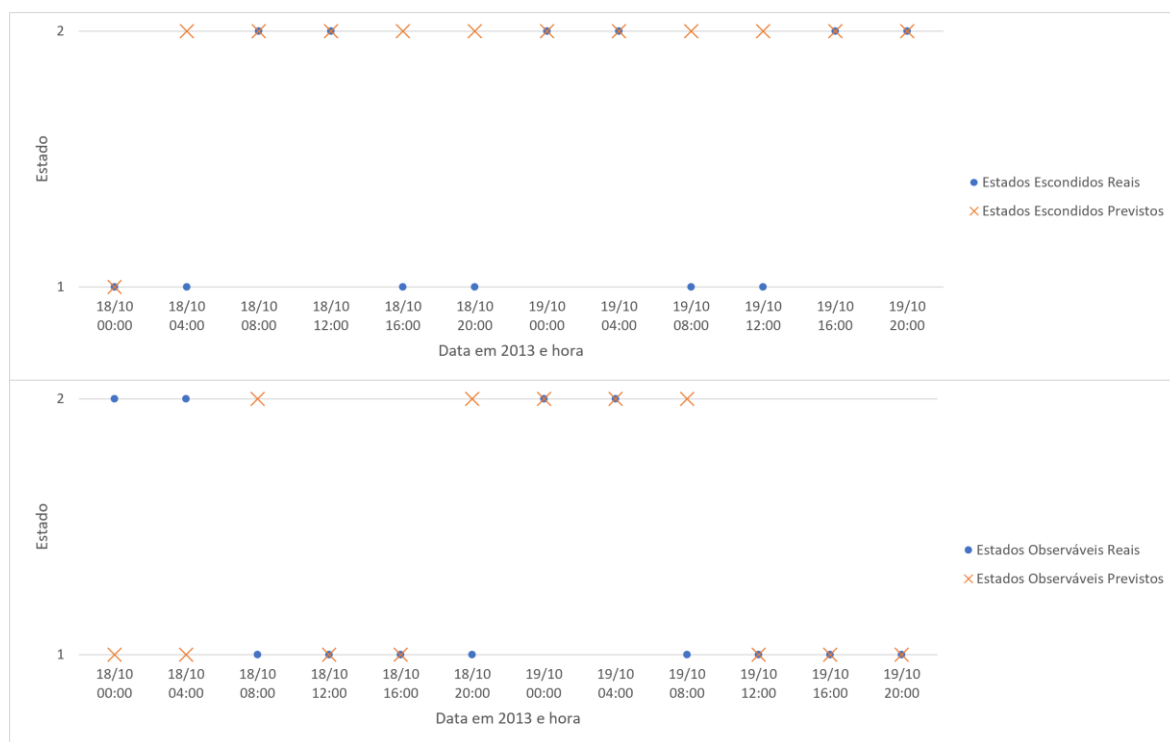


Figura 24 - Comparação dos estados escondidos e estados observáveis no teste 4

#### **4.6. TESTE 5 – 2 ESTADOS ESCONDIDOS E 2 ESTADOS OBSERVÁVEIS COM PADRÃO DA TELEVISÃO**

Para um novo teste, tentou-se encontrar um padrão nos dados para se poder calcular as probabilidades necessárias para o HMM.

Inicialmente, pensou-se em tentar encontrar um padrão no consumo agregado para averiguar se existia um padrão nesse consumo em relação às horas de medição. Pela análise efetuada, os picos no consumo são um acontecimento raro, e das vezes que aconteceram não foi possível determinar um padrão entre eles para se apurar uma possível razão deles estarem a acontecer.

Como ainda era necessário tentar encontrar um padrão para realizar este teste, verificou-se cada equipamento para ver se havia algum padrão em algum deles. O equipamento que pareceu mais provável ter um foi a televisão. Normalmente, as pessoas ligam a televisão num certo padrão, como por exemplo quando chegam do trabalho, enquanto cozinham, de manhã ou à noite, enquanto que os outros equipamentos não pareciam ter um padrão tão regular como a televisão.

Analisou-se as 2097 medições para procurar um padrão para o consumo da televisão. Foi possível notar-se imediatamente que a televisão ou tinha ou não tinha consumo, logo é possível assumir que a televisão apenas consumia energia quando estava ligada. Ao analisar os dados, notou-se que as horas que as pessoas desta residência consumiam mais televisão era entre as 08:00 e as 09:00, e também entre as 16:00 e as 02:00.

Para criar os estados escondidos, escolheu-se como estado 1 o período em que as pessoas mais consumiam televisão como indicado anteriormente, e para o estado 2 as restantes horas. Como a televisão só consome quando está ligada, o estado observável 1 é quando há consumo de energia pelo equipamento 8, enquanto o estado 2 é quando não há nenhum consumo. As suas percentagens, podem ser observadas na figura 25.

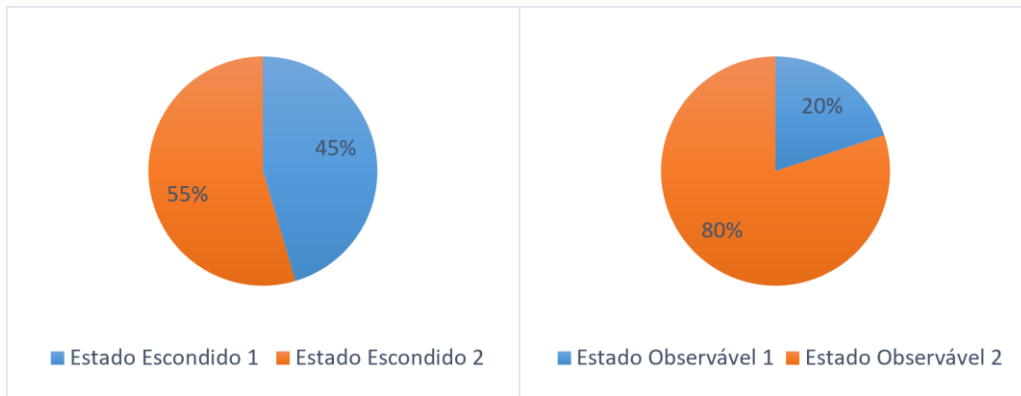


Figura 25 - Número dos estados escondidos e observáveis do teste 5

Pelos testes anteriores, verificou-se que o livraria “HMM” funciona melhor quando uma menor quantidade de dados, por isso utilizou-se 5967 medições, que são 20 dias, para calcular as probabilidades necessárias. As probabilidades de transição são apresentadas na figura 26 e as de emissão na figura 27.

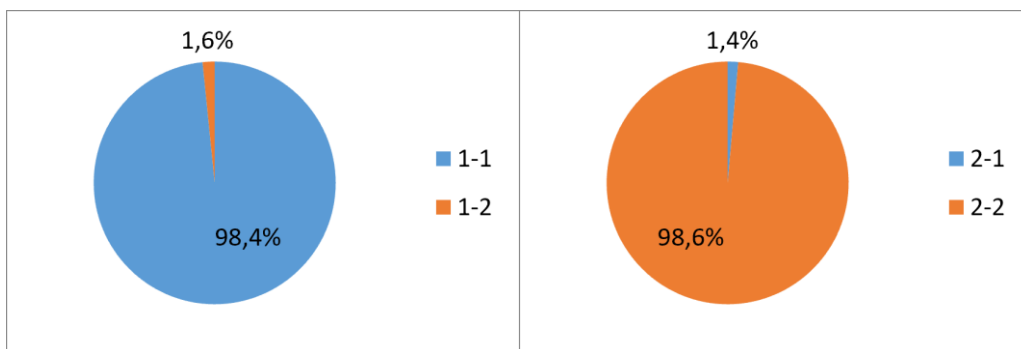


Figura 26 - Probabilidade de transição para teste 5

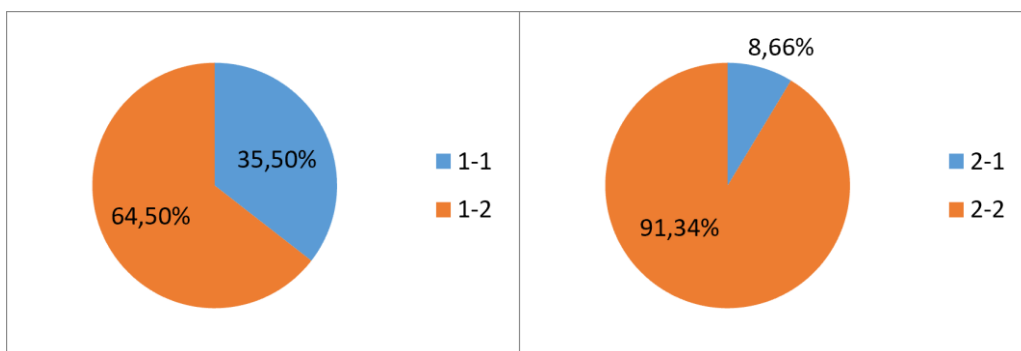


Figura 27 - Probabilidade de emissão para teste 5

Para o treino do HMM, usou-se as observações dos 20 dias, para depois prever para os três dias seguintes, que é um treino de 85% e teste de 15%.

Como é possível observar pela figura 28, os resultados foram muito melhores para a previsão dos estados observáveis, tendo em conta a quantidade de dados, mas os resultados para a previsão dos estados escondidos foram desafortunados, porque em comparação com os testes anteriores não foi um dos melhores resultados obtidos.

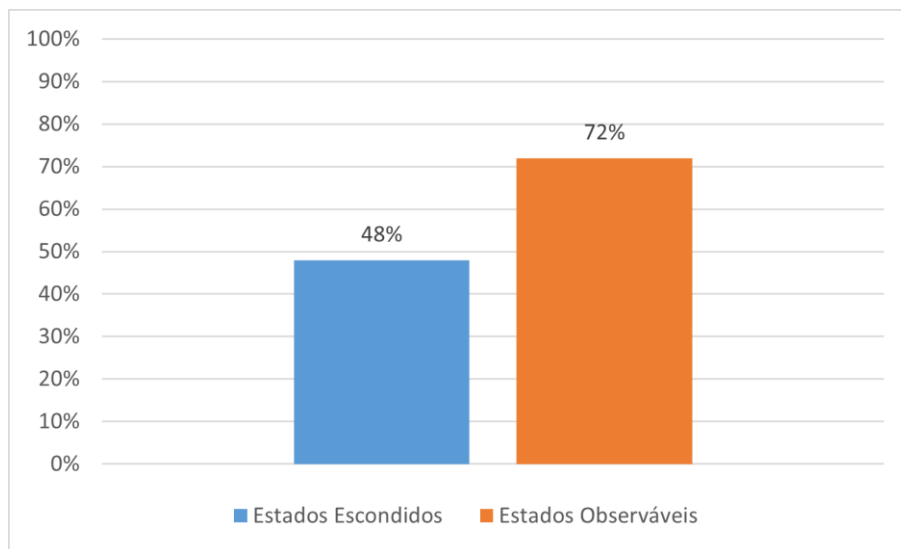


Figura 28 - Resultados do teste 5

O possível motivo do bom resultado da previsão dos estados observáveis é por ter previamente encontrado um padrão para poder calcular as probabilidades em função disso. Como o padrão encontrado, apesar de não acontecer rigorosamente todos os dias, tem uma ocorrência alta, o HMM conseguiu prever corretamente a maior parte dos dados.

Na figura 29 é possível comparar os estados escondidos e observáveis dos dados reais e dos previstos pelo HMM.

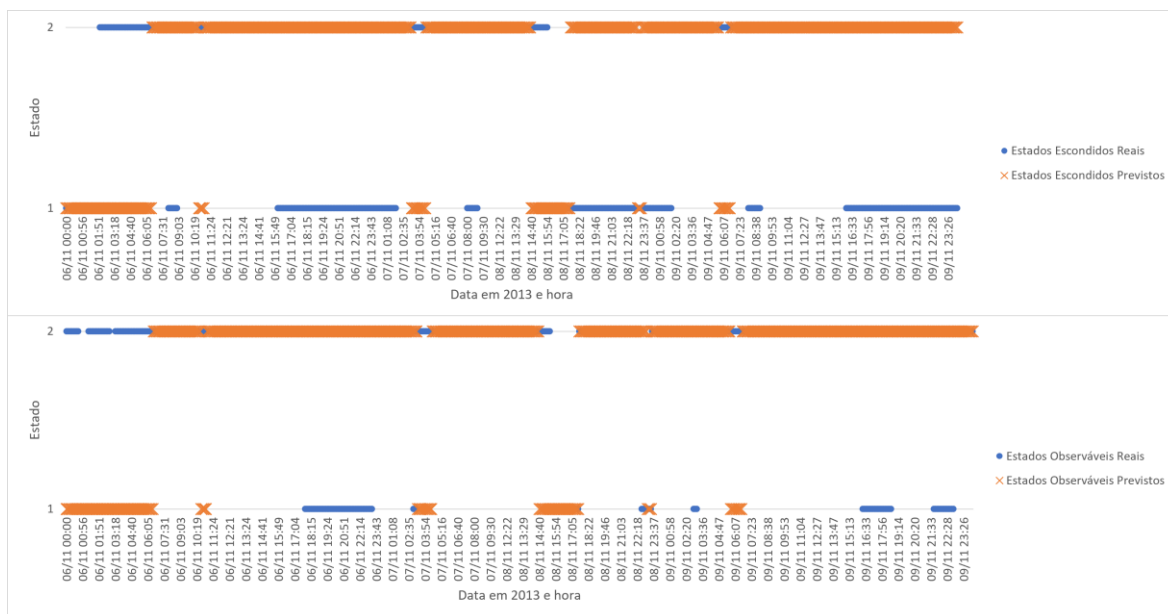


Figura 29 - Comparação dos estados escondidos e observáveis no teste 5

#### 4.7. DISCUSSÃO DOS RESULTADOS

Na tabela 3 estão presentes os resultados obtidos após a realização do treino *Viterbi* de todos os testes realizados.

Tabela 3 - Resumo dos resultados

Teste 1	Estados Escondidos	Estados Observáveis
Teste 1	37%	28%
Teste 1 (70%-30%)	28%	25%
Teste 2	36%	20%
Teste 3	42%	51%
Teste 4	58%	58%
Teste 5	48%	72%

No teste 1, os 4 estados escondidos foram definidos, dependendo da agregação dos consumos, enquanto que os 9 estados observáveis são os equipamentos que estão a ser

medidos. Ao realizar o primeiro teste, o HMM foi capaz de prever corretamente 37% dos estados escondidos e 28% dos estados observáveis. Também foi feito o mesmo teste, mas para 70% de treino e 30% de previsão, em que o HMM foi capaz de prever corretamente 28% dos estados escondidos e 25% dos estados observáveis. Com esses valores, decidiu-se realizar mais testes com outros HMMs para tentar obter melhores valores.

No teste 2 foi feita uma experiência em que se utilizou exatamente o mesmo número de estados observáveis e escondidos do teste 1, mas tentou-se dividir os estados de forma a cada um ser 25% cada. Foi possível assim ter melhores resultados.

Para o teste seguinte, o teste 3, tentou-se fazer uma abordagem mais simples, para testar o programa, de forma a verificar se o problema estava na quantidade de dados que se estava a utilizar. Por causa disso, foi apenas definido dois estados escondidos e dois estados observáveis. Com esse HMM, obteve-se melhores resultados que nos testes anteriores, tendo previsto corretamente 42% dos estados escondidos e 51% dos estados observáveis. Foi possível notar-se que se obteve resultados melhores, logo uma razão plausível para isso é o facto de o programa funcionar melhor com uma menor quantidade de dados.

Para o teste 4, simplificou-se ainda mais os dados, manteve-se os mesmos estados do teste anterior, mas aumentou-se a diversidade de dados para mais dias, reduzindo assim a quantidade de dados. Ao juntar a maior diversidade de dados para treino e a menor quantidade de dados que o programa teve de usar para treinar o HMM, os resultados foram melhores que os do teste 3, tendo previsto corretamente 28% dos estados escondidos e observáveis.

Como os testes 3 e 4 tiveram melhores resultados, decidiu-se que o teste 5 também teria apenas 2 estados escondidos e observáveis. Mas desta vez, procurou-se previamente um padrão para definir as probabilidades em função disso. Dessa forma, obteve-se os melhores resultados de todos os testes na previsão dos estados observáveis, tendo previsto corretamente 72%. A nível dos estados escondidos, os resultados pioraram um bocão, tendo apenas previsto corretamente 48% dos estados escondidos, ficando só atrás do teste 4.

O modelo realizado nos testes não é o mais apropriado para lidar com dados temporais, uma vez que é baseado em probabilidades. No caso do teste 5, a probabilidade de o modelo estar no estado de televisão ligada, terá sempre a mesma probabilidade de se mudar ou se manter no estado, independentemente de estar ou não mais próximo de mudar.



# 5. CONCLUSÕES

Nos últimos anos, tem havido um aumento no consumo de energia e na produção da mesma por fontes de geração renovável. Por isso, é necessário conseguir diminuir a energia consumida ou alocar alguma dessa energia a horas sem pico, de forma a diminuir a penetração de energia na rede durante as horas de pico. Uma outra opção é conseguir realizar mudanças de consumo de maneira a seguir as variações da geração renovável, como por exemplo aumentar o consumo quando há maior geração e diminuir quando há menos.

Por causa disso, a previsão de padrões de consumo em edifícios residências é um aspecto importante, pois é uma forma de se conseguir desenvolver métodos de identificação de consumidores para distribuir o consumo de energia para diferentes horários para diminuir a quantidade de consumo.

Após uma análise de diversos modelos alternativos, decidiu-se implementar este trabalho utilizando o *Hidden Markov Model*, uma vez que em diversos artigos, os autores conseguiram realizar testes de previsão de consumo em edifícios residenciais e obtiveram bons resultados.

Na fase de testes, escolheu-se usar uma livreria de HMM no R, *HMM*, para desenvolver um HMM que fosse capaz de prever da melhor forma possível o consumo de um edifício residencial. A base de dados escolhida foi a base de dados RETIF, que contém muitas medições ao longo de vários meses, no qual especifica os equipamentos que estão a ser contados para as medições e o número de pessoas a morar naquela habitação. Foram realizados 5 testes para tentar criar um HMM que conseguisse prever o consumo da habitação 1 da base de dados. Ao longo dos testes realizados, obteve-se diversos resultados, sendo alguns deles menos bons que outros. De todos os testes, foi o teste 5 que teve os melhores resultados. A principal causa é a de já se ter encontrado um padrão previamente e dessa forma, foi possível calcular as probabilidades de transição e de emissão, os estados escondidos e observáveis com base nesse padrão. Nos outros testes, não foi feito um estudo de um padrão inicialmente.

Também foi possível verificar que os resultados foram melhores quando se diminuiu o número de dados que o programa tinha de analisar. Uma causa provável desse acontecimento é o programa não estar desenvolvido para lidar com uma quantidade elevada de dados para análise do HMM e para o seu treino. Uma razão pelos resultados serem melhores com uma quantidade de dados menores, é de o HMM ter menos chances de errar ao realizar o treino do HMM e ao fazer a previsão. Por isso, para o último teste, foi utilizada uma quantidade maior de dados, comparativamente aos testes 3 e 4, para se poder verificar se os resultados eram realmente fiáveis e foi verificado que os resultados foram melhores que os encontrados nos testes anteriores com uma menor quantidade de dados.

Apesar de no último teste ter-se tido um resultado bom, não foi um dos melhores que estão presentes na literatura para casos semelhantes. Uma possível causa é o da livraria usada não ter sido completamente personalizada para a base de dados e o problema em questão. Dos artigos lidos, os autores desenvolveram os seus próprios programas de HMM para a sua realização. Logo, se se tivesse desenvolvido um novo programa, poderia ter sido possível obter melhores resultados, mas devido às limitações em programação, não foi possível realizar com sucesso um novo programa implementando os HMM de raiz.

Como trabalho futuro propõe-se o desenvolvimento de um modelo HMM criado de forma personalizada para o problema em questão, além da exploração de formas alternativas de se lidar com os padrões temporais, de forma a ultrapassar as limitações verificadas pelos HMM devido à representação das mudanças de estado através de probabilidades, perdendo-se a componente temporal e contextual.

## Referências Bibliográficas

- Abreu, J. M., Câmara Pereira, F., & Ferrão, P. (2012). Using pattern recognition to identify habitual behavior in residential electricity consumption. *Energy and Buildings*, 49, 479–487. <https://doi.org/10.1016/j.enbuild.2012.02.044>
- Abu Al-Haija, Q. (2021). A Stochastic Estimation Framework for Yearly Evolution of Worldwide Electricity Consumption. *Forecasting*, 3(2), 256–266. <https://doi.org/10.3390/forecast3020016>
- Allahverdyan, A., & Galstyan, A. (2011). Comparative analysis of Viterbi Training and Maximum Likelihood estimation for HMMs. *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011*.
- Bock, T. (2021). *What are the Strengths and Weaknesses of Hierarchical Clustering?* <https://www.displayr.com/strengths-weaknesses-hierarchical-clustering/>
- Burlando, M. (2018). *Hidden Markov Models — Part 1: the Likelihood Problem*. [https://medium.com/@Ayra\\_Lux/hidden-markov-models-part-1-the-likelihood-problem-8dd1066a784e](https://medium.com/@Ayra_Lux/hidden-markov-models-part-1-the-likelihood-problem-8dd1066a784e)
- Burlando, M. (2019). *Hidden Markov Models — Part 2: the Decoding Problem*. [https://medium.com/@Ayra\\_Lux/hidden-markov-models-part-2-the-decoding-problem-c628ba474e69](https://medium.com/@Ayra_Lux/hidden-markov-models-part-2-the-decoding-problem-c628ba474e69)
- Burns, N. (2019). *Pattern Recognition via Principal Components Analysis*. <https://www.sqlservercentral.com/articles/pattern-recognition-via-principal-components-analysis>
- Chai, A. (2018). Household consumption patterns and sectoral composition of growing economies: a review of the interlinkages. *Griffith University*, 02(1), 1–59.
- Crandall, A. S., & Cook, D. J. (2010). Using a Hidden Markov Model for resident identification. *Proceedings - 2010 6th International Conference on Intelligent Environments, IE 2010*, 74–79. <https://doi.org/10.1109/IE.2010.21>
- Edwards, R. E., New, J., & Parker, L. E. (2012). Predicting future hourly residential electrical consumption: A machine learning case study. *Energy and Buildings*, 49, 591–

603. <https://doi.org/10.1016/j.enbuild.2012.03.010>
- Eisner, J. (2002). *An interactive spreadsheet for teaching the forward-backward algorithm*. *July*, 10–18. <https://doi.org/10.3115/1118108.1118110>
- Fan, S., & Chen, L. (2006). Short-term load forecasting based on an adaptive hybrid method. *IEEE Transactions on Power Systems*, *21*(1), 392–401. <https://doi.org/10.1109/TPWRS.2005.860944>
- Fang, H., Zhang, Y., Liu, M., & Shen, W. (2018). Clustering and Analysis of Household Power Load Based on HMM and Multi-factors. *Proceedings of the 2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design, CSCWD 2018*, 455–460. <https://doi.org/10.1109/CSCWD.2018.8465353>
- Fett, D., Keles, D., Kaschub, T., & Fichtner, W. (2019). Impacts of self-generation and self-consumption on German household electricity prices. *Journal of Business Economics*, *89*(7), 867–891. <https://doi.org/10.1007/s11573-019-00936-3>
- Figueiredo, V., Rodrigues, F., Vale, Z., & Gouveia, J. B. (2005). An electric energy consumer characterization framework based on data mining techniques. *IEEE Transactions on Power Systems*, *20*(2), 596–602. <https://doi.org/10.1109/TPWRS.2005.846234>
- Flores-Quiroz, A., Palma-Behnke, R., Zakeri, G., & Moreno, R. (2016). A column generation approach for solving generation expansion planning problems with high renewable energy penetration. *Electric Power Systems Research*, *136*, 232–241. <https://doi.org/https://doi.org/10.1016/j.epsr.2016.02.011>
- Germano, T. (1999). *Self Organizing Maps*. <http://davis.wpi.edu/~matt/courses/soms/>
- Guenounou, A., Aillerie, M., Mahrane, A., Bouzaki, M., Boulouma, S., & Charles, J. P. (2021). Human home daily living activities recognition based on a LabVIEW implemented hidden Markov model. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-021-10814-2>
- Gutierrez-Osuna, R. (2002). Pattern analysis for machine olfaction: a review. *IEEE Sensors Journal*, *2*(3), 189–202. <https://doi.org/10.1109/JSEN.2002.800688>
- Himmelmann, L. (2015). Package “HMM” Title HMM -Hidden Markov Models. *Proceedings of the IEEE*, *77*(2), 257–286.
- Hossain, K. A. (2012). Global Energy Consumption Pattern and GDP. *International Journal of Renewable Energy Technology Research*, *1*(1), 23–29.

- Iliopoulos, N., Esteban, M., & Kudo, S. (2020). Assessing the willingness of residential electricity consumers to adopt demand side management and distributed energy resources: A case study on the Japanese market. *Energy Policy*, *137*, 111169. <https://doi.org/https://doi.org/10.1016/j.enpol.2019.111169>
- Jozi, A., Pinto, T., Praça, I., & Vale, Z. (2019). Decision Support Application for Energy Consumption Forecasting. *Applied Sciences*, *9*(4). <https://doi.org/10.3390/app9040699>
- Jurafsky, D., & Martin, J. H. (2020). Hidden Markov Models. *Statistics in Human Genetics and Molecular Biology*, 151–162. <https://doi.org/10.1201/b16380-13>
- Kumar, N. (2019). *Advantages and Disadvantages of Principal Component Analysis in Machine Learning*. [http://theprofessionalspoint.blogspot.com/2019/03/advantages-and-disadvantages-of\\_4.html](http://theprofessionalspoint.blogspot.com/2019/03/advantages-and-disadvantages-of_4.html)
- Lim, K. Z., Lim, K. H., Wee, X. Bin, Li, Y., & Wang, X. (2020). Optimal allocation of energy storage and solar photovoltaic systems with residential demand scheduling. *Applied Energy*, *269*, 115116. <https://doi.org/https://doi.org/10.1016/j.apenergy.2020.115116>
- Lin, G., Pan, F., Yang, Y., Yang, L., He, G., & Fan, S. (2018). The Pattern Recognition of Residential Power Consumption Based on HMM. *2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 413–418. <https://doi.org/10.1109/ISGT-Asia.2018.8467905>
- Lobo, V. (2007). *Mapas Auto-Organizados - SOM SOM Kohonen 's Self- Organising Maps O que é um SOM ? Perspectiva histórica Visão geral do SOM*. 1–9.
- Mi, L., Xu, T., Sun, Y., Yang, H., Wang, B., Gan, X., & Qiao, L. (2021). Promoting differentiated energy savings: Analysis of the psychological motivation of households with different energy consumption levels. *Energy*, *218*, 119563. <https://doi.org/https://doi.org/10.1016/j.energy.2020.119563>
- Miljković, D. (2017). Brief review of self-organizing maps. *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1061–1066. <https://doi.org/10.23919/MIPRO.2017.7973581>
- Mokhov, V., Eltsova, M., & Bauer, M. (2021). Regional differentiation of energy consumption in urban households. *Journal of Physics: Conference Series*, *1886*(1). <https://doi.org/10.1088/1742-6596/1886/1/012014>

- Murray, D., Stankovic, L., & Stankovic, V. (2017). An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. *Scientific Data*, 4(1), 160122. <https://doi.org/10.1038/sdata.2016.122>
- Nikolaou, T. G., Kolokotsa, D. S., Stavrakakis, G. S., & Skias, I. D. (2012). On the Application of Clustering Techniques for Office Buildings' Energy and Thermal Comfort Classification. *IEEE Transactions on Smart Grid*, 3(4), 2196–2210. <https://doi.org/10.1109/TSG.2012.2215059>
- Park, S., Serpedin, E., & Qaraqe, K. (2013). Gaussian Assumption: The Least Favorable but the Most Useful [Lecture Notes]. *IEEE Signal Processing Magazine*, 30(3), 183–186. <https://doi.org/10.1109/MSP.2013.2238691>
- Pinto, T., Faia, R., Navarro-Caceres, M., Santos, G., Corchado, J. M., & Vale, Z. (2018). Multi-Agent-Based CBR Recommender System for Intelligent Energy Management in Buildings. *IEEE Systems Journal*, 13(1), 1084–1095. <https://doi.org/10.1109/JSYST.2018.2876933>
- Pinto, T., Praça, I., Vale, Z., & Silva, J. (2021). Ensemble learning for electricity consumption forecasting in office buildings. *Neurocomputing*, 423, 747–755. <https://doi.org/https://doi.org/10.1016/j.neucom.2020.02.124>
- Ramanathan, N. (2006). *Applications of Hidden Markov Models*. University of Maryland. <http://www.cs.umd.edu/~djacobs/CMSC828/ApplicationsHMMs.pdf>
- Rosin, A., Hõimoja, H., Möller, T., & Lehtla, M. (2010). Residential electricity consumption and loads pattern analysis. *Proceedings of the 2010 Electric Power Quality and Supply Reliability Conference*, 111–116. <https://doi.org/10.1109/PQ.2010.5550009>
- Santos, G., Pinto, T., Vale, Z., Morais, H., & Praca, I. (2012). Balancing market integration in MASCEM electricity market simulator. *IEEE Power and Energy Society General Meeting*, 1–8. <https://doi.org/10.1109/PESGM.2012.6345652>
- Su, R., McBride, E. C., & Goulias, K. G. (2020). Pattern recognition of daily activity patterns using human mobility motifs and sequence analysis. *Transportation Research Part C: Emerging Technologies*, 120, 102796. <https://doi.org/https://doi.org/10.1016/j.trc.2020.102796>
- Sütterlin, B., Brunner, T. A., & Siegrist, M. (2011). Who puts the most energy into energy conservation? A segmentation of energy consumers based on energy-related behavioral characteristics. *Energy Policy*, 39(12), 8137–8152.

<https://doi.org/10.1016/j.enpol.2011.10.008>

- Ullah, I., Ahmad, R., & Kim, D. (2018). A Prediction Mechanism of Energy Consumption in Residential Buildings Using Hidden Markov Model. *Energies*, *11*(2). <https://doi.org/10.3390/en11020358>
- Vale, Z., Faria, P., Abrishambaf, O., Gomes, L., & Pinto, T. (2021). MARTINE—A Platform for Real-Time Energy Management in Smart Grids. *Energies*, *14*(7). <https://doi.org/10.3390/en14071820>
- Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., Han, M., & Zhao, X. (2018). A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews*, *82*, 1027–1047. <https://doi.org/https://doi.org/10.1016/j.rser.2017.09.108>
- Wu, S., Zheng, X., You, C., & Wei, C. (2019). Household energy consumption in rural China: Historical development, present pattern and policy implication. *Journal of Cleaner Production*, *211*, 981–991. <https://doi.org/10.1016/j.jclepro.2018.11.265>