

Sensibilidade de políticas cientes de carbono para VMs: efeitos da especificação, do cenário e da comparabilidade da carga

Guilherme Fernandes Moraes da Silva¹, Daniel Cordeiro¹

¹Escola de Artes, Ciências e Humanidades – Universidade de São Paulo (USP)
São Paulo, SP – Brasil

{guilherme.fernandes01,daniel.cordeiro}@usp.br

Abstract. *The sensitivity of comparisons among 11 virtual machine (VM) allocation and migration policies is evaluated across 30 configurations and 6,820 seed-paired runs. First, comparing two ordering specifications revealed changes in VM placement and metrics. Second, scenarios alter contrast magnitudes and policy orderings; under counterfactual inverted PUE, the minimum-energy policy produced up to $3.2\times$ the estimated emissions of the minimum-emission policy. Third, in 10 of the 26 distinct configurations, admission differs across policies and three emission metrics yield incompatible rankings, precluding a single ranking. Thus, conclusions about a policy's carbon performance jointly depend on its specification, the experimental scenario, and the comparability of the admitted workload. These elements should therefore be incorporated into the design of comparative evaluations of carbon-aware policies.*

Resumo. *Avalia-se a sensibilidade da comparação entre 11 políticas de alocação e migração de máquinas virtuais (VMs) em 30 configurações, com 6.820 execuções pareadas por semente. Primeiro, a comparação entre duas especificações de ordenação mostrou alterações no posicionamento das VMs e nas métricas resultantes. Segundo, o cenário altera magnitudes dos contrastes e ordenações: sob PUE invertido contrafactual, a política de menor consumo de energia resultou em até $3,2\times$ as emissões estimadas da política de menor emissão. Terceiro, em 10 das 26 configurações distintas, a admissão difere entre políticas e três métricas de emissão produzem ordenações incompatíveis, impedindo uma ordenação escalar. Assim, conclusões sobre o desempenho em carbono de uma política dependem conjuntamente de sua especificação, do cenário experimental e da comparabilidade da carga admitida. Esses elementos devem, portanto, integrar o desenho de avaliações comparativas de políticas cientes de carbono.*

1. Introdução

Data centers respondem por cerca de 1% do consumo mundial de eletricidade [Masanet et al. 2020]. Em infraestruturas geodistribuídas, o impacto climático depende também da intensidade de carbono da eletricidade, que varia entre regiões e ao longo do tempo [Vasconcelos et al. 2023]. Essa heterogeneidade motiva estratégias que deslocam cargas entre regiões ou períodos de menor intensidade [Vasconcelos et al. 2022, Wiesner et al. 2021, Acun et al. 2023]. Neste trabalho, o foco são políticas que posicionam e migram máquinas virtuais (VMs) entre infraestruturas geodistribuídas.

Uma política ciente de carbono não opera isoladamente: suas decisões dependem da infraestrutura disponível, da intensidade de carbono, da rede que sustenta as migrações e de detalhes da própria especificação. Regras aparentemente secundárias fazem parte da política tanto quanto o critério declarado: quando dois *hosts* ou dois *data centers* empatam sob esse critério, é a regra de desempate que decide o posicionamento, de modo que uma especificação que não defina os desempates não determina completamente o comportamento da política. Além disso, minimizar energia não é o mesmo que minimizar emissões, e os dois objetivos podem induzir ordenações distintas entre políticas [Hanafy et al. 2023]. Por fim, quando as políticas admitem quantidades distintas de carga, a ordenação passa a depender de como a emissão se relaciona à carga admitida, e essa escolha deixa de ser um detalhe de apresentação. Alterações nesses elementos podem mudar não apenas a magnitude das emissões estimadas, mas também o posicionamento das VMs e a ordenação entre as políticas.

Neste trabalho, investiga-se até que ponto as conclusões comparativas permanecem estáveis quando esses elementos são variados de forma controlada. São comparadas 11 políticas de alocação e migração em 30 configurações e 6.820 execuções, conduzidas com uma única versão fixa do simulador e com regras de desempate declaradas. O desenho é pareado por semente.

As contribuições são quatro. (i) Uma avaliação reprodutível de 11 políticas em 30 configurações, executada com uma única versão fixa do simulador e com as regras determinísticas de desempate explicitadas como parte da especificação. (ii) Uma avaliação de como os contrastes observados e as ordenações variam com o perfil de intensidade de carbono, a rede, a modificação conjunta de topologia e capacidade e a atribuição de PUE. (iii) Uma análise explícita de comparabilidade que mostra que, em 10 das 26 configurações distintas, a ordenação entre políticas depende da métrica de emissão adotada; nessas condições, não se declara uma ordenação escalar. (iv) Um protocolo de avaliação em três etapas que exige especificar a política, variar controladamente o cenário e verificar a comparabilidade da carga admitida antes de interpretar as ordenações.

2. Trabalhos relacionados

A base conceitual das políticas cientes de carbono avaliadas aqui está nos estudos sobre abordagens *follow-the-renewables* em multinuvens [Vasconcelos et al. 2022], no gerenciamento de recursos sensível ao preço e à fonte de energia em *data centers* geodistribuídos [Ali et al. 2021] e no gerenciamento de energia renovável e de pegada de carbono em ambientes multinuvem [Xu and Buyya 2020]. Três das políticas avaliadas são adaptações dessas propostas, ajustadas ao arcabouço aqui utilizado: *cnemesis* e *followme_s*, que migram VMs em tempo de execução, e *wsnb*, que é uma política estática, atuando apenas na alocação inicial. A outra política com migração, *follow_renewables*, é uma heurística reativa implementada especificamente no arcabouço utilizado neste estudo. Os nomes são mantidos para identificar a proposta de origem de cada adaptação; nenhuma comparação numérica deste estudo é atribuída às implementações originais. O comportamento implementado de cada política é descrito na Seção 3.2 (Tabela 3).

No eixo temporal, [Wiesner et al. 2021] quantificam o potencial do deslocamento temporal da carga para reduzir as emissões na nuvem. A respeito de infraestrutura, o Carbon Explorer [Acun et al. 2023] trata o projeto de *data centers* cientes de carbono

de forma holística, em contraste com a avaliação comparativa de políticas em simulação conduzida aqui. A tensão entre otimizar energia e otimizar carbono é formulada por [Hanafy et al. 2023]. Quanto à estimativa de emissões, [Saraiva et al. 2025] implementam um *plugin* de pegada de carbono para SimGrid/Batsim e confrontam as emissões estimadas em simulação com execuções reais instrumentadas; o problema tratado aqui é complementar e distinto: dada uma estimativa de emissão, sob que condições uma ordenação entre políticas é interpretável.

Em conjunto, esses trabalhos investigam mecanismos relacionados em recortes distintos. [Vasconcelos et al. 2022], por exemplo, avaliam uma única topologia de nove *data centers* e realizam uma execução por configuração. O recorte aqui adotado combina a especificação operacional da política, a variação controlada do cenário e a verificação da comparabilidade da carga admitida como condição para interpretar uma ordenação entre políticas. A Tabela 1 compara a unidade de decisão, o mecanismo e as dimensões avaliadas por quatro desses trabalhos.

Tabela 1. Trabalhos relacionados: unidade de decisão, mecanismos e dimensões avaliadas nas propostas originais.

Estudo	Unidade e mecanismo	Dimensões da avaliação
[Xu and Buyya 2020]	Requisições; deslocamento e brownout	Energia não renovável, carbono e tempo de resposta
[Ali et al. 2021]	VMs; alocação e migração	Energia, tempo de execução e custos
[Vasconcelos et al. 2022]	VMs; migração entre nuvens	Congestionamento, migração e energia não renovável
[Hanafy et al. 2023]	<i>Jobs</i> ; deslocamento temporal/espacial e ajuste de recursos	Energia e carbono sob diferentes graus de flexibilidade
Este estudo	VMs; alocação e migração	Especificação, cenário e comparabilidade da carga

3. Metodologia

O estudo compara políticas sob o mesmo modelo energético e de carbono, seguindo três passos:

1. especificar explicitamente a política, incluindo regras de desempate e ordens de enumeração que influenciem a decisão;
2. variar controladamente os fatores do cenário sob cargas pareadas;
3. verificar a comparabilidade da carga admitida antes de interpretar a ordenação.

3.1. Plataforma, modelo de energia e semântica de migração

Os experimentos foram conduzidos em um simulador de eventos discretos próprio, construído sobre o SimGrid 4.1.1 [Casanova et al. 2025], no qual as políticas de alocação e migração são módulos intercambiáveis que observam o mesmo estado do *cluster*. O cenário base tem nove *data centers* geodistribuídos e 450 *hosts*, 50 por *data center* (DC), distribuídos em cada DC em três classes: $25 \times (42.560 \text{ MIPS}, 64 \text{ GiB}, 35\text{--}50 \text{ W})$,

$15 \times (85.120 \text{ MIPS}, 128 \text{ GiB}, 70\text{--}100 \text{ W})$ e $10 \times (170.240 \text{ MIPS}, 256 \text{ GiB}, 140\text{--}200 \text{ W})$. A capacidade de CPU é expressa em MIPS, convenção adotada neste cenário; para execução no SimGrid, esses valores são mapeados para FLOP/s, a unidade nativa do simulador, à razão de 1 MFLOP/s por MIPS. A potência instantânea (P_h) é interpolada linearmente entre o repouso e o pico em função do uso da CPU. Cada execução simula 14.400 s (4 h) de operação, mesmo horizonte nas 30 configurações, correspondente a dois tempos de vida médios de VM e contendo todas as chegadas dos prefixos de *trace* utilizados, com alocação decidida a cada passo de 60 s sobre as VMs chegadas nesse passo, pontos de decisão de migração a cada 600 s, enlaces inter-DC de 1 Gb/s (latência derivada da distância geográfica: 0,5 ms por grau, mínimo de 1 ms) e interfaces de *host* de 10 Gb/s. Os nove DCs simulados são associados às localidades de Paris, São Paulo, Virgínia, Seul, Singapura, Dubai, Canberra, Pune e Joanesburgo. Como o horizonte de 4 h cobre apenas parte do ciclo diurno dos perfis de intensidade de carbono, as comparações se restringem a essa janela temporal.

As cargas provêm de entradas derivadas do Azure Packing Trace 2020 (solicitações de VMs) e do Google ClusterData 2011 (*jobs* compostos de *tasks*) [Hadary et al. 2020, Wilkes 2011]; aqui, VM designa a entidade simulada. Os arquivos de entrada foram obtidos de [Vasconcelos et al. 2022]; neste estudo, alterou-se apenas o instante de término declarado na linha de controle final de cada arquivo, e os arquivos resultantes foram redistribuídos no artefato deste trabalho. Cada linha de criação de VM fornece o instante de chegada, o identificador, o número de núcleos (usado apenas na instanciação da VM) e um valor numérico que, no simulador original, parametrizava a duração da execução. Neste cenário, esse valor é reinterpretado como demanda em MIPS, restrita ao intervalo $[1; 170.000]$, limite definido na configuração do cenário. As VMs são ordenadas por (chegada, identificador), e cada configuração usa o prefixo das primeiras n VMs nessa ordenação. No *trace* do Google, as chegadas são reescaladas para uma janela de 7.200 s. RAM de 8.192 MiB (8 GiB), imagem de 8.192 MiB e taxa de geração de páginas sujas de 40 Mb/s não vêm do *trace* e são impostas pela configuração. Os tempos de vida são sorteados de uma exponencial de média 7.200 s por um gerador pseudoaleatório inicializado com a semente da execução; essa escolha é sintética e não preserva as vidas dos rastros públicos. VMs cujo tempo de vida ultrapassa o horizonte permanecem ativas até o fim da janela simulada. A semente controla exclusivamente os tempos de vida; as decisões das políticas são determinísticas dado o estado. A Tabela 2 caracteriza os prefixos utilizados; a correspondência registro a registro com os rastros públicos completos não foi recuperada.

Tabela 2. Prefixos transformados utilizados. d_i : demanda nominal em MIPS, derivada de duração, não CPU medida; P95 por interpolação linear. Pico: chegadas em 60 s. As estatísticas não descrevem os rastros públicos completos.

Entrada / pressão	n	Chegadas (min)	d_i : mediana / P95	Pico
Azure / moderada	2.000	5–10	531,7 / 3440,4	1847
Azure / alta	10.000	5–80	869,8 / 6685,0	1847
Google / moderada	2.000	0–120	430,3 / 27258,2	577
Google / alta	9.000	0–120	458,7 / 29621,7	577

A energia estimada por *host* h em um intervalo Δt em que a utilização é, portanto,

P_h , permanece constante é $E_h^{\text{kWh}}(\Delta t) = P_h [\text{W}] \cdot \Delta t [\text{s}] / (3,6 \times 10^6)$; a energia da instalação do *data center* d é $E_d^{\text{inst}}(\Delta t) = \text{PUE}_d \sum_{h \in d} E_h^{\text{kWh}}(\Delta t)$; e a emissão operacional estimada do *data center* é $C_d(\Delta t) = E_d^{\text{inst}}(\Delta t) \cdot I_d(t)$, em que o divisor $3,6 \times 10^6$ converte joules em quilowatt-hora, PUE_d , entre 1,17 e 1,405 conforme o DC no cenário base, multiplica a energia de TI, e $I_d(t)$ é a intensidade de carbono regional horária do *data center* d , mantida constante em cada passo de 60 s, em $\text{gCO}_2\text{e/kWh}$. A energia reportada e usada nas ordenações é a energia da instalação, isto é, após a aplicação do PUE. PUE (*Power Usage Effectiveness*) é a razão entre energia total e de TI; não mede trabalho útil nem emissões isoladamente. As séries históricas usam a intensidade do ciclo de vida da geração elétrica. Não se contabilizam fabricação dos equipamentos, construção ou descarte físico.

A emissão é reportada em gCO_2e e o volume migrado em MiB. A migração é contabilizada no momento em que a decisão é emitida (migrações emitidas); a transferência pode não ser concluída na janela simulada. VMs para as quais nenhum *host* com capacidade é encontrado são descartadas, sem fila de espera. A taxa de admissão é definida como a fração das VMs oferecidas que encontram alocação. A migração usa o modelo de *pre-copy* [Clark et al. 2005] do SimGrid, com dois parâmetros fixos: o conjunto de trabalho (*working set*) de cada VM igual a 0,9 vezes a sua RAM e a intensidade de páginas sujas (*dirty page intensity*) igual à razão entre a taxa de páginas sujas da VM e a banda da interface de rede do *host* de criação, saturada no intervalo $[0, 1]$; a geração de páginas sujas depende da atividade, não ocorrendo na ociosidade. Rejeição e destruição não recebem custo energético próprio, assim como a cópia de memória e os equipamentos de rede; sua magnitude omitida é desconhecida.

3.2. Políticas e especificação determinística

Foram comparadas 11 políticas: oito atuam apenas na alocação inicial e três também emitem migrações em tempo de execução (*cnemesis*, *follow_renewables* e *followme_s*); *wsnb* é implementada nesta avaliação apenas como política de alocação inicial.

Na tabela, LCD, FR, FMS e CN abreviam *lowest_carbon_dc*, *follow_renewables*, *followme_s* e *cnemesis*. WSNB usa FNV-1a de 64 bits do *vm_id*, módulo o número de DCs, sobre a lista ordenada por *dc_id*. Em CN, a condição $0,9I_{\text{receptor}} < I_{\text{remetente}}$ é sempre satisfeita pela partição da mediana neste desenho; não garante economia de 10%.

As seis políticas genéricas não são cientes de carbono nem emitem migrações: *first_fit* usa o primeiro *host* viável na ordem de enumeração do cenário e *round_robin* percorre essa mesma ordem a partir de um cursor persistente, que só avança quando há colocação. *best_fit* e *worst_fit* escolhem o *host* viável de menor e de maior folga residual de CPU (capacidade menos demanda alocada, com a RAM como filtro binário de viabilidade), respectivamente. *ffd* e *bfd* ordenam as VMs do lote de chegada do passo corrente por demanda decrescente antes de aplicar First Fit e Best Fit. Os limiares e tetos da Tabela 3 permanecem fixos em todas as configurações, sem ajuste por cenário.

Como *followme_s* exclui apenas o *host* de origem da lista de destinos, a política pode migrar dentro do mesmo DC; assim, nem todas as suas migrações são inter-DC.

Tabela 3. Comportamento implementado das cinco políticas cientes de carbono. BF: Best Fit; I : intensidade instantânea em gCO₂e/kWh. Os tetos contam decisões por intervalo.

Política	Alocação inicial	Migração e limites
LCD	DC viável de menor I ; BF interno	Não migra
WSNB	Origem por hash do ID; mantém se $I \leq 400$ e couber; senão, como LCD	Não migra
FR	Como LCD	Evacua DCs com $I > 400$ para DC viável de menor $I \leq 400$, BF interno; uma decisão por VM, sem teto global
FMS	Como LCD	Hosts com utilização nominal $< 0,20$; destino por I e BF, inclusive intra-DC; teto 50
CN	Como LCD (previsão de horizonte zero)	Remetentes acima da mediana de I ; BF global nos receptores; demais DCs consolidam localmente ($< 0,20$); teto conjunto 20

Eventos são ordenados deterministicamente por instante, tipo (Submit antes de Terminate no mesmo instante) e `vm_id`. Empates entre VMs usam demanda decrescente, chegada crescente e `vm_id` crescente; entre *data centers*, intensidade de carbono crescente e `dc_id` crescente; empates de folga entre *hosts* são resolvidos pelo primeiro *host* na ordem de enumeração do cenário.

Políticas que varrem *hosts* usam a ordem de enumeração do cenário, definida pela ordem de declaração dos *data centers* no arquivo de configuração e pela ordem lexicográfica dos *hosts* dentro de cada DC. As decisões de alocação e migração usam o mesmo estado, antes da aplicação dos lotes do passo corrente. Assim, tanto os desempates quanto a ordem de enumeração fazem parte da especificação reproduzível.

3.3. Desenho experimental

A avaliação compreende 30 configurações de cenário e 6.820 execuções das 11 políticas, todas sob a mesma versão fixa do simulador. Em 4 pares de configurações, a única diferença é a troca do perfil sintético de junho pelo de janeiro; uma configuração de cada par é redundante por construção, pois o perfil sintético não modela sazonalidade e repete a mesma curva diurna ao longo do ano, restando 26 configurações não redundantes.

As 16 configurações do cenário base foram executadas com $n = 30$ sementes e as demais 14 com $n = 10$, com uma execução por combinação política-semente. As 16 configurações do cenário base formam um fatorial $2 \times 4 \times 2$: duas cargas derivadas de *traces* públicos (Azure e Google), quatro perfis sintéticos de intensidade (junho, janeiro, uniforme e anticorrelacionado) e dois níveis de pressão sobre a infraestrutura, com 2.000 VMs no nível moderado e 10.000 (Azure) ou 9.000 (Google) no alto.

São cinco os perfis de intensidade de carbono. No cenário base, os perfis sintéticos de junho (`real_jun`) e de janeiro (`real_jan`) são curvas diurnas ajustadas às médias anuais de cada região reportadas por [Vasconcelos et al. 2023].

Tabela 4. Famílias do desenho: carbono uniforme e anticorrelação são perfis do cenário base; as demais famílias aplicam as modificações indicadas, mantidas as demais condições. Em `geo3` e `net100`, dois fatores variam conjuntamente.

Família	Característica / modificação
Carbono uniforme (<code>flat</code>)	Perfil de intensidade constante (475 gCO ₂ e/kWh)
Anticorrelação (<code>anticorr</code>)	Perfil com fase do ciclo diurno de intensidade distinta por DC (valores no artefato)
Rede restrita (<code>net100</code>)	Enlaces inter-DC: banda 1 Gb/s → 100 Mb/s e latência ×5
Topologia reduzida (<code>geo3</code>)	DCs (9 → 3) e capacidade (450 → 150 <i>hosts</i>)
Capacidade heterogênea (<code>hetcap</code>)	Redistribuição dos <i>hosts</i> por DC, de 20 (menor intensidade média) a 80 (maior intensidade média), crescentes com o posto; total de 450 preservado
PUE invertido (<code>pue</code>)	Reatribuição do PUE por DC em ordem inversa à intensidade média, de 1,12 a 1,50 (menor intensidade média → maior PUE)
Electricity Maps (<code>em2021</code>)	Perfil sintético substituído por série histórica (2021-06-01, 00:00–04:00 UTC)

Entre os perfis do cenário base, `flat` usa intensidade uniforme, que não diferencia destinos, e `anticorr` testa a resposta das políticas a intensidades temporalmente defasadas entre regiões, sem representar uma rede elétrica realista. As famílias, incluindo as extensões estruturais e históricas, são resumidas na Tabela 4.

As séries do Electricity Maps [Electricity Maps 2026] foram obtidas por meio de licença acadêmica, o que veda a redistribuição; o artefato disponibiliza o procedimento de obtenção e a janela utilizada. As 14 configurações adicionais distribuem-se entre `net100` (4), `geo3` (2), `hetcap` (2), `pue` (2) e `em2021` (4). Excluindo as quatro que usam séries históricas, restam 22 configurações não redundantes baseadas apenas em perfis sintéticos. Em `geo3` e em `net100`, dois fatores variam juntos e ficam confundidos, e não podem ser separados neste desenho.

3.4. Análise estatística e comparabilidade

Cada semente determina uma realização da carga oferecida, compartilhada por todas as políticas, de modo que as comparações sejam pareadas. Comparações entre configurações com números diferentes de sementes usam apenas as dez sementes comuns às duas. A análise principal compara as três políticas com migração à política estática ciente de carbono `lowest_carbon_dc` pelo teste de Wilcoxon com p exato, seguido do método de Holm-Bonferroni dos três valores- p em cada configuração; o controle do erro é, portanto, por configuração. As comparações entre todos os 55 pares das 11 políticas são exploratórias. Diferenças nulas são descartadas; a interpretação como deslocamento de localização requer simetria das diferenças.

As variações percentuais são calculadas como $100 \cdot \bar{d}/\bar{x}_{\text{ref}}$, em que $\bar{d}$ é a média das diferenças pareadas (política menos referência) e $\bar{x}_{\text{ref}}$ é a média da referência; valores negativos indicam emissão menor. Esse resumo descritivo não é o estimando do teste.

A concordância entre ordenações é resumida pelo τ_b de Kendall. A equivalência

de emissões totais é avaliada pelo procedimento de dois testes unilaterais (TOST) baseado em Wilcoxon, com margem operacional pré-especificada de $\pm 1\%$ da média amostral de emissão da política de referência em cada configuração; como a margem depende dos dados, verificou-se que nenhuma decisão de equivalência muda sob perturbação de ± 2 erros-padrão dessa média. A margem não representa um limiar de relevância prática definido pelo domínio.

O contraste descritivo entre grupos compara os mínimos únicos de emissão das políticas cientes e não cientes de carbono. Os mínimos de emissão e energia em cada configuração são definidos pela média entre sementes.

Como VMs não alocadas são descartadas, avalia-se também a comparabilidade da carga admitida. São consideradas três formas de expressar a emissão: emissão total (N_0), emissão por VM alocada (N_1) e emissão dividida pela soma da demanda em MIPS das VMs admitidas (N_2), com razões por execução antes da média. Nenhuma é neutra: N_0 combina emissão e quantidade de carga admitida, enquanto N_1 e N_2 introduzem denominadores que também variam entre políticas. As configurações são classificadas em três grupos: admissão integral, admissão desigual com ordenação estável sob as três métricas e ordenação sensível à métrica. A classificação é um diagnóstico de robustez, não uma escolha de métrica preferida; admissão integral não demonstra execução equivalente.

Para distinguir admissão de execução, a reanálise posterior usa quatro execuções instrumentadas compatíveis com a campanha (Tabela 6). Configurações, decisões e métricas científicas coincidem com as execuções originais. Define-se $S = \sum_{i \in A} (W_i - R_i)$, em que A é o conjunto de VMs admitidas, $W_i = 2Hd_i 10^6$ FLOPs sintéticos, $H = 14.400$ s e R_i restante antes da destruição ou do horizonte (zero se concluída). Suspensão não acrescenta progresso. Reportam-se ainda $Q = S / \sum_{i \in O} W_i$, com O sendo o conjunto de VMs oferecidas, e C/S ; não medem utilidade da aplicação nem equivalência de serviço.

3.5. Reprodutibilidade

As 6.820 execuções usam a mesma versão fixa do simulador, com *commit* e *checksum* SHA-256 declarados nos artefatos. Os arquivos do pacote de artefatos foram verificados por *checksums*, sem divergência, e 46 execuções selecionadas para verificação de determinismo reproduziram byte a byte o registro de decisões e todas as colunas de métricas utilizadas na análise, exceto a coluna de tempo de relógio, que não é determinista. Código, configurações, entradas derivadas e ferramentas de análise estão publicamente disponíveis¹. As limitações de construto, instrumentação e reprodutibilidade estão na Seção 5.

4. Resultados

4.1. Sensibilidade à especificação da política

A especificação de uma política não termina no critério de escolha: as regras de desempate declaradas na Seção 3.2 tornam o posicionamento determinístico, mas condicionado a essas ordens, e *wsnb* depende ainda da função de *hash* que mapeia o identificador da VM ao DC de origem (Tabela 3).

¹<https://github.com/GuilhermeFernandes01/sscad-2026/tree/sscad-2026>

A sensibilidade à especificação da ordenação foi avaliada comparando a especificação anterior, que ordenava apenas pela chave principal e não especificava a ordem relativa entre elementos equivalentes, com a final, de chaves secundárias determinísticas, sobre o mesmo conjunto de execuções da família histórica: as sequências divergem nos 440 pares, primeiro pelo ID da VM em 428 e pelo destino em 12 (11 políticas $\times$ 4 configurações $\times$ 10 sementes). Nos 12 cenários base distintos e dez sementes comuns, os 1.320 pares também divergem; a política de menor emissão muda em quatro. A mediana de $|100(C_{\text{final}}/C_{\text{anterior}} - 1)|$ por execução é 2,16% na carga moderada e 0,22% na alta; *cnemesis* chega a +10,92% na média por cenário. A comparação mostra onde a divergência começa e como se propaga, sem isolar o desempate: também mudam o hash de origem de *wsnb* e a ordem de soma das métricas.

Avaliou-se também, pela sequência de colocações no registro de decisões, se as três políticas com migração produzem o mesmo posicionamento que *lowest_carbon_dc* sob a especificação final: nas 22 configurações sintéticas não redundantes e nas dez sementes comuns, são 660 pares (política com migração $\times$ configuração $\times$ semente), com posicionamento idêntico em 340 e divergente nos 320 restantes. Essa comparação não isola o efeito do desempate e não é usada para atribuir causalidade. A identidade observada é condicional. Dos 660 pares, em 140 a política com migração não emitiu nenhuma migração, e todos os 140 apresentaram alocação idêntica à de *lowest_carbon_dc*; para *follow_renewables*, a identidade valeu em 160 dos 220 pares. Sob intensidade uniforme de 475, nenhum destino satisfaz o limiar $I \leq 400$ de *follow_renewables*, impedindo sua migração.

4.2. Sensibilidade ao cenário

No cenário contrafactual de PUE invertido (Tabela 5), escolher a política de menor consumo de energia resultou em 18,95 kgCO₂e estimados (*best_fit*), contra 5,84 (*followme_s*), diferença de 13,11 kg; a admissão integral não prova execução equivalente. Mesmo nas 12 configurações não redundantes do cenário base, energia e carbono selecionaram políticas distintas em duas, ambas sob o perfil anticorrelacionado.

A Tabela 5 sintetiza as modificações de cenário analisadas a seguir, com o escopo de interpretação ao lado de cada resultado. No cenário contrafactual de PUE invertido, escolher a política de menor consumo de energia multiplicou as emissões em relação à de menor emissão de carbono. Mesmo nas 12 configurações não redundantes do cenário base, energia e carbono selecionaram políticas distintas em duas, ambas sob o perfil anticorrelacionado.

Na configuração de rede restrita, as emissões totais das três políticas com migração ficaram dentro da margem operacional de $\pm 1\%$ de *lowest_carbon_dc*. O TOST sustenta equivalência apenas para a emissão total: como a carga admitida difere entre as políticas e a ordenação é sensível à métrica de emissão, esse resultado não caracteriza equivalência entre as políticas.

Nas configurações *geo3*, o contraste entre os mínimos de emissão dos dois grupos foi atenuado nas duas cargas (Tabela 5); o número de *data centers* e a capacidade total diminuem conjuntamente, portanto o resultado não identifica efeito causal do número de *data centers*. Nessa família ocorreu o único contraste do par *cnemesis* $\times$ *lowest_carbon_dc* com sinal favorável à política estática de referência (+3,12%),

Tabela 5. Síntese da sensibilidade da comparação às modificações de cenário. Cada linha preserva a métrica natural da análise; grandezas de linhas distintas não são comparáveis entre si.

Modificação	Resultado principal	Escopo da interpretação
PUE invertido	Emissão ao minimizar energia: até $3,246 \times$ o mínimo de emissão	Contrafactual; máximo sob admissão integral; outra configuração com 98,75% a 100% de admissão
Rede restrita	FR/FMS/CN vs. LCD: $-0,03\%$ a $-0,14\%$ nas emissões totais; TOST $\pm 1\%$	Azure, alta, junho sintético; admissão de 94,82% a 96,58%; não demonstra equivalência de serviço
geo3	Contraste relativo ao mínimo genérico: 58,37% $\rightarrow$ 2,90% (Azure); 65,89% $\rightarrow$ 6,40% (Google)	Dez sementes comuns; DCs e capacidade variam juntos; Google com admissão desigual e ordenação sensível à métrica
Capacidade heterogênea	Emissão/base em junho sintético: LCD +101,54%; genéricas de +11,86% a +23,06%	Admissão integral; em anticorrelação, a emissão de LCD diminui 15,02%
Séries históricas	Direção preservada em 10/12 contrastes; a política de menor emissão muda em 1/4 configurações	Azure com admissão integral; Google entre 98,75% e 100%; janela histórica única de quatro horas

no Google moderado, em configuração sensível à métrica de emissão; ele é significativo na sua família de três valores- p , mas, por resultar de varredura das 22 configurações não redundantes sintéticas, é exploratório. Sob capacidade heterogênea e junho sintético, o aumento da emissão de `lowest_carbon_dc` foi maior do que a faixa observada nas heurísticas genéricas.

Com séries históricas (Electricity Maps), os 12 contrastes comparam as três políticas com migração a `lowest_carbon_dc` nas quatro configurações históricas (Tabela 5); nos dois contrastes sem preservação de direção não houve inversão de sinal: a diferença no lado sintético era nula (empate entre as duas políticas), e a troca da política de menor emissão total opõe `followme_s`, sob o perfil sintético de junho, a `follow_renewables`, sob as séries históricas. No Google, a ordenação permanece estável às três métricas apesar da admissão desigual (Tabela 5).

4.3. Sensibilidade à definição do trabalho comparado

Quando as políticas não admitem a mesma quantidade de carga, a emissão total passa a refletir simultaneamente a política e a carga admitida. Das 26 configurações distintas, 10 tiveram admissão integral, 6 apresentaram admissão desigual com ordenação estável às três formas de expressão avaliadas e 10 apresentaram ordenação sensível à métrica de emissão (11, 7 e 12 das 30, contadas as 4 redundantes). Nas primeiras, as políticas admitem as mesmas VMs, mas o trabalho executado pode diferir. Nas seguintes, a ordenação só é interpretada em conjunto com as taxas de admissão. Nas restantes, nenhuma ordenação escalar é defensável.

A Figura 1 mostra, para duas dessas configurações, o posto das onze políticas sob as três métricas da Seção 3.4. A diferença entre a maior e a menor demanda admitida é de

44,73% e 51,76% do máximo, respectivamente. Nessas configurações, portanto, adotar uma única ordenação exige escolher previamente uma definição do trabalho comparado que os dados não determinam.

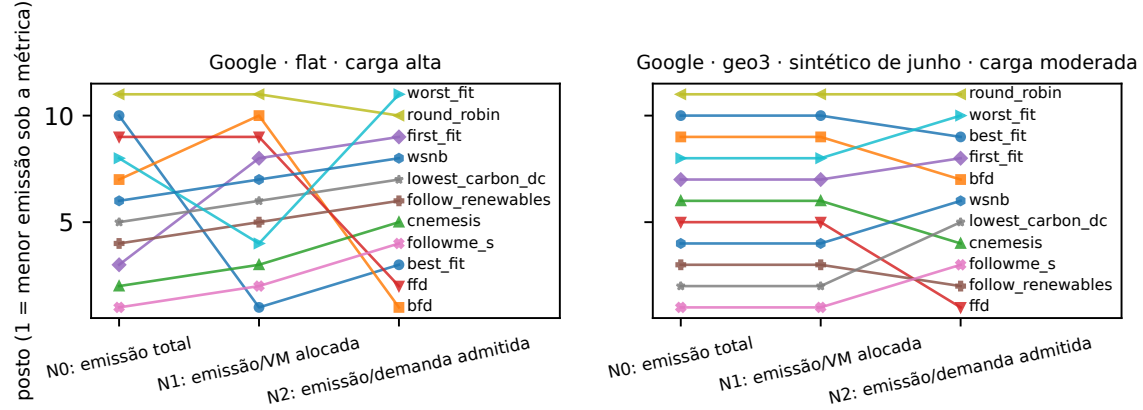


Figura 1. Ordenação das políticas sob N_0 , N_1 e N_2 em duas configurações com admissão desigual de carga. Posto 1 indica o menor valor; empates exatos recebem posto médio. Razões calculadas por execução antes da média. Políticas na Seção 3.2.

A oferta comum é de 704,8 PF sintéticos (Tabela 6). *follow_renewables* admite mais VMs que *lowest_carbon_dc*, mas executa 0,12% menos trabalho; suas emissões estimadas são 4,69% menores e C/S , 4,57% menor. Isso não demonstra serviço equivalente nem superioridade nas demais sementes.

Tabela 6. Emissões e serviço: Azure, junho sintético, carga alta, semente 42. a : admissão; S : trabalho executado; Q : fração executada do trabalho oferecido; $PF=10^{15}$ FLOPs sintéticos; C/S em gCO_2e/PF . Siglas na Tabela 3; uma observação por política.

Política	E (kWh)	$\hat{C}$ (kgCO ₂ e)	a (%)	S (PF)	Q (%)	$\hat{C}/S$ (g/PF)
LCD	143,9	60,71	94,94	473,2	67,14	128,3
FMS	142,9	60,12	95,45	475,5	67,46	126,4
FR	139,8	57,86	95,64	472,6	67,06	122,4
CN	143,8	60,61	95,59	477,0	67,68	127,1

5. Ameaças à validade

A principal ameaça é à validade de construto. Nenhuma das métricas N_0 , N_1 e N_2 mede o trabalho efetivamente executado: a análise de progresso acrescentada cobre apenas quatro execuções e conta FLOPs sintéticos, não utilidade entregue. As contagens de migração representam decisões emitidas; o contador de volume pode sobrestimar blocos interrompidos e omitir progresso no horizonte, não constituindo limite inferior. O modelo de energia não inclui um termo próprio para os equipamentos de rede e as magnitudes absolutas não foram calibradas externamente; compartilhar o modelo não elimina vieses diferenciais entre políticas.

Internamente, as configurações *geo3* combinam topologia e capacidade instalada, e as de rede restrita, banda e latência; seus efeitos não podem ser atribuídos isoladamente

a um dos fatores. A alocação em lotes de 60 s pode limitar a generalização a sistemas que decidem a cada chegada. Usar o identificador da VM como chave final de desempate torna a execução reproduzível, mas impõe uma ordem arbitrária entre alternativas equivalentes.

Com $n = 10$, a família exploratória dos 55 pares não pode atingir o nível de 5% após o método de Holm, quaisquer que sejam os dados; a ausência de significância reflete a resolução do desenho, não evidência de ausência de efeito. As sementes representam realizações dos tempos de vida sorteados, e não amostras de *data centers*, redes elétricas ou políticas. A análise foi reproduzida internamente com o mesmo código e ambiente: isso evidencia integridade, mas não constitui implementação independente nem exclui erros sistemáticos compartilhados. Externamente, as cargas provêm de dois *traces* públicos tomados por prefixo, o que não constitui amostragem aleatória. Não se avaliou sensibilidade à média das vidas, RAM, páginas sujas ou aos limiares e tetos fixos das políticas. Os perfis de intensidade de carbono são majoritariamente sintéticos; a série histórica cobre uma única janela de 4 h sem sazonalidade, e não se verificou se as médias dos perfis sintéticos seguem o mesmo escopo de contabilização (ciclo de vida) das séries históricas. Todos os resultados provêm de uma única plataforma de simulação, com um único modelo de energia, migração e rede. VMs concluídas permanecem alocadas e expirações no passo da criação podem ser ignoradas, afetando a ocupação; o efeito sobre a ordenação não foi isolado. As conclusões restringem-se às configurações efetivamente executadas.

6. Conclusão e trabalhos futuros

A comparação entre as duas especificações de ordenação mostrou alterações em posicionamentos e métricas; por isso, desempates e ordens de enumeração devem compor a especificação da política. Mudanças no perfil de intensidade de carbono, no PUE, na capacidade e nas configurações `geo3` e de rede restrita alteraram contrastes e ordenações. A interpretação também depende da comparabilidade da carga admitida: em 10 das 26 configurações distintas, as três métricas de emissão induzem ordenações incompatíveis e, como nenhuma mede o trabalho executado, os dados não sustentam uma única ordenação escalar. A análise de quatro execuções não demonstra equivalência de serviço na campanha. A especificação da política, o cenário e a definição do trabalho comparado devem, portanto, ser tratados como parte do desenho experimental: omitir qualquer um desses elementos pode transformar uma propriedade do desenho em uma aparente propriedade da política. Trabalhos futuros incluem isolar os fatores confundidos, ampliar as janelas históricas e a instrumentação do trabalho executado pelas VMs.

Este trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP #2023/00811-0), do CCD Cidades Carbono Neutro (FAPESP #2024/01115-0) e do grupo de pesquisa THUS (Techno-Human Systems of the Future) do Centro Internacional de Pesquisa (IRC) ‘Transitions’ (CNRS, USP e FAPESP CIP #2025/01171-0).

A ferramenta Codex (OpenAI) foi utilizada como apoio à revisão do manuscrito e à formatação do texto, incluindo tabelas e figuras. Os autores permanecem integralmente responsáveis pelo conteúdo do artigo.

Referências

- [Acun et al. 2023] Acun, B., Lee, B., Kazhamiaka, F., Maeng, K., Gupta, U., Chakkavarthy, M., Brooks, D., and Wu, C.-J. (2023). Carbon explorer: A holistic framework for designing carbon aware datacenters. In *ASPLOS 2023: Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*, pages 118–132. ACM.
- [Ali et al. 2021] Ali, H., Zakarya, M., Rahman, I. U., Khan, A. A., and Buyya, R. (2021). Followme@ls: Electricity price and source aware resource management in geographically distributed heterogeneous datacenters. *Journal of Systems and Software*, 175:110907.
- [Casanova et al. 2025] Casanova, H., Giersch, A., Legrand, A., Quinson, M., and Suter, F. (2025). Lowering entry barriers to developing custom simulators of distributed applications and platforms with simgrid. *Parallel Computing*, 123:103125.
- [Clark et al. 2005] Clark, C., Fraser, K., Hand, S., Hansen, J. G., Jul, E., Limpach, C., Pratt, I., and Warfield, A. (2005). Live migration of virtual machines. In *NSDI'05: Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation - Volume 2*, pages 273–286. USENIX Association.
- [Electricity Maps 2026] Electricity Maps (2026). AE, AU-NSW, BR-CS, FR, IN-WE, KR, SG, US-MIDA-PJM and ZA 2021 Hourly Carbon Intensity Data (Version February 10, 2026). Electricity Maps. <https://www.electricitymaps.com>.
- [Hadary et al. 2020] Hadary, O., Marshall, L., Menache, I., Pan, A., Greeff, E. E., Dion, D., Dorminey, S., Joshi, S., Chen, Y., Russinovich, M., and Moscibroda, T. (2020). Protean: VM allocation service at scale. In *14th USENIX Symposium on Operating Systems Design and Implementation (OSDI 20)*, pages 845–861. USENIX Association.
- [Hanafy et al. 2023] Hanafy, W. A., Bostandoost, R., Bashir, N., Irwin, D., Hajiesmaili, M., and Shenoy, P. (2023). The war of the efficiencies: Understanding the tension between carbon and energy optimization. In *HotCarbon '23: Proceedings of the 2nd Workshop on Sustainable Computer Systems*, pages 1–7. ACM.
- [Masanet et al. 2020] Masanet, E., Shehabi, A., Lei, N., Smith, S., and Koomey, J. (2020). Recalibrating global data center energy-use estimates. *Science*, 367(6481):984–986.
- [Saraiva et al. 2025] Saraiva, G., Vasconcelos, M., Bruschi, S., Carastan-Santos, D., and Cordeiro, D. (2025). Estimating CO₂ emissions of distributed applications and platforms with SimGrid/Batsim. In *Anais do XXVI Simpósio em Sistemas Computacionais de Alto Desempenho*, pages 109–120, Porto Alegre, RS, Brasil. SBC.
- [Vasconcelos et al. 2023] Vasconcelos, M., Cordeiro, D., Da Costa, G., Dufossé, F., Nicod, J.-M., and Rehn-Sonigo, V. (2023). Optimal sizing of a globally distributed low carbon cloud federation. In *2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, pages 203–215. IEEE.
- [Vasconcelos et al. 2022] Vasconcelos, M. F. S., Cordeiro, D., and Dufossé, F. (2022). Indirect network impact on the energy consumption in multi-clouds for follow-the-renewables approaches. In *Proceedings of the 11th International Conference on Smart Cities and Green ICT Systems (SMARTGREENS)*, pages 44–55.

- [Wiesner et al. 2021] Wiesner, P., Behnke, I., Scheinert, D., Gontarska, K., and Thamsen, L. (2021). Let’s wait awhile: how temporal workload shifting can reduce carbon emissions in the cloud. In *Middleware ’21: Proceedings of the 22nd International Middleware Conference*, pages 260–272. ACM.
- [Wilkes 2011] Wilkes, J. (2011). Google cluster-usage traces: format + schema. Technical report, Google Inc.
- [Xu and Buyya 2020] Xu, M. and Buyya, R. (2020). Managing renewable energy and carbon footprint in multi-cloud computing environments. *Journal of Parallel and Distributed Computing*, 135:191–202.